Data, Metrics and Monitoring in CGIAR - a strategic study

Other

Data, metrics and monitoring in CGIAR – a strategic study

December 2014

CGIAR

Independent Science and Partnership Council

Cover image: courtesy Neil Palmer/CIAT

The ISPC encourages appropriate use of its published information, including quotation, but asks that the source be acknowledged. The recommended reference for this publication is:

ISPC. 2014. *Data, metrics and monitoring in CGIAR – a strategic study*. Rome, Italy, CGIAR Independent Science and Partnership Council (ISPC). 88 pp.

Contents

Acronyms and abbreviations

A4NH	Agriculture for Nutrition and Health (CRP)
AAFC	Agriculture & Agri-Food Canada
AAS	Aquatic Agricultural Systems (CRP)
ACIAR	Australian Centre for International Agricultural Research
AKIS	agricultural knowledge and innovation systems
CCAFS	Climate Change, Agriculture and Food Systems (CRP)
CED	chronic energy deficiency
CIARD	Coherence in Information for Agricultural Research for Development
CIAT	International Center for Tropical Agriculture
CIMMYT	International Maize and Wheat Improvement Center
CMEF	Common Monitoring and Evaluation Framework
CRP	CGIAR Research Program
CSI	Consortium for Spatial Information
DALY	disability-adjusted life year
DFID	Department for International Development, United Kingdom
DHS	Demographic and Health Survey
DOI	digital object identifier
EC	European Community
EEA	European Environment Agency
EU	European Union
FADN	Farm Accountancy Data Network
FaNSI	FAO Food and Nutrition Security Index
FAO	Food and Agriculture Organization of the United Nations
FTA	Forests, Trees and Agroforestry (CRP)
GDP	gross domestic product
GECAFS	Global Environmental Change and Food Systems
GFAR	Global Forum on Agricultural Research
GIS	geographic information system
GODAN	Global Open Data for Agriculture and Nutrition
GPS	global positioning system
GRiSP	Global Rice Science Partnership (CRP)
ICAR	Indian Council of Agricultural Research
ICARDA	International Center for Agricultural Research in the Dry Areas
ICRAF	World Agroforestry Centre
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
ICT	information and communications technology
ICT-KM	Information and Communications Technology – Knowledge Management (CGIAR)
IDO	intermediate development outcomes
IFAD	International Fund for Agricultural Development
IFPRI	International Food Policy Research Institute
IITA	International Institute of Tropical Agriculture
ILRI	International Livestock Research Institute
IRRI	International Rice Research Institute
ISPC	Independent Science and Partnership Council
IUCN	International Union for Conservation of Nature
L&F	Livestock & Fish (CRP)
LSMS	Living Standards Measurement Study/Studies (World Bank)
LSMS-ISA	LSMS Integrated Surveys on Agriculture
NAHARP	National Agri-Environmental Health Analysis and Reporting Program
NARS	national agricultural research system(s)
NGO	nongovernmental organization
NRM	natural resources management
PESTEL	political, economic,social, technological, environmental and legal
PIM	Policy, Institutions & Markets (CRP)
OECD	Organisation for Economic Co-operation and Development
OGP	Open Government Partnership
R&D	research and development
RCT	randomized controlled trial
RSB	Roundtable on Sustainable Biofuels
RSPO	Roundtable for Sustainable Palm Oil
RTB	Roots, Tubers and Bananas (CRP)
RTRS	Roundtable on Responsible Soy
SAN	Sustainable Agriculture Network

SDG	Sustainable Development Goal
SDSN	Sustainable Development Solutions Network
SIAC	Strengthening Impact Assessment in CGIAR (a program led by SPIA)
SINGER	System-wide Information Network for Genetic Resources
SLO	system-level outcome
SMART	specific, measurable, attainable, realistic and time-bound
SMS	Short Message Service
SPIA	Standing Panel on Impact Assessment
SRF	Strategy and Results Framework
SRTM elevation	spatial database on elevation
SSA	Sub-Saharan Africa
TFP	total factor productivity
UN	United Nations
UNDP	United Nations Development Programme
UNEP	United Nations Environment Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
USAID	United States Agency for International Development
USDA	United States Department of Agriculture
VDSA	Village Dynamics in South Asia
WAW	World Agriculture Watch
WCMC	World Conservation Monitoring Centre
WEAI	Women's Empowerment in Agriculture Index
WHO	World Health Organization
WLE	Water, Land and Ecosystems (CRP)

I.

ISPC Synthesis and Commentary

Background and rationale of the study

The requirement for progress and agreement on metrics, indicators and data management has emerged as an urgent challenge presented by the changing agenda of CGIAR and the broadening of its system-level outcomes (SLOs). The need for new concepts, tools and procedures for program and system-level analysis has been reinforced by the ISPC's review of the 15 CRPs, and a number of previous ISPC studies (Social Sciences, ISPC, 2009; Natural Resources Management, ISPC, 2012a). These have all led to concern that CGIAR has not had a strong record of systematically measuring, monitoring and curating data on the principal systems that it is seeking to influence. Long-term data sets on changes in the resource base, agricultural production and the livelihoods of target beneficiaries of research have not been maintained. The need for a set of standard metrics was also highlighted as a priority in the new performance management system being developed by the Consortium for CRP monitoring and evaluation and for impact assessment. CGIAR needs to undertake appropriate target setting and develop the means to monitor progress toward the achievement of intermediate development outcomes (IDOs) and the overall CGIAR impacts.

An ideal system of common metrics will serve three main purposes in CGIAR:

1. to measure changes in agricultural productivity (including across scales, from field to regional and global) and to monitor associated impacts on the environment, environmental services, livelihoods and other dimensions of human welfare;
2. to assess/measure causal linkages and trade-offs among IDOs and SLOs;
3. to provide comparability and common understanding that allows assessment of the CGIAR portfolio and reporting, both in relation to specific CRPs and in an aggregated manner at the system level.

Although there is agreement among CGIAR partners and stakeholders about the need for new metrics and data management systems, there is still debate about what to measure, how, where and by whom. However, the general conclusion emerging from the CGIAR Science Forum 2011, held in Beijing (ISPC, 2011), was that CGIAR seems to have comparative advantages in developing protocols and standards for monitoring key performance variables of agricultural systems, and in characterizing cropping and crop–livestock systems, and developing metrics for key environmental variables.

The general guiding principles for the strategic study on metrics planned by the ISPC were to: (i) promote the science for improving metrics within CGIAR in a transparent, consistent, reproducible, robust and unifying manner; (ii) ensure that the metric systems adopted are relevant across disciplines and spatial scales to integrate different IDOs (e.g. productivity and human well-being and/or environmental sustainability); (iii) strengthen the community of practice working on metrics appropriate for CGIAR; (iv) identify a set of simple, low-cost metrics and decision tools that can be used for CGIAR's specific needs; and (v) identify and prioritize new research to fill the gaps and develop the science underlying metrics (ISPC, 2013a).

The study was undertaken by a panel of external experts chaired by Dr Ken Giller of Wageningen University, the Netherlands. It was designed to address the Terms of Reference described in an ISPC concept note (ISPC, 2013a). The study included a survey questionnaire to the CRPs about their planned use of metrics and indicators for CRP- and system-level IDOs, and a workshop with CRP members, selected partners, donors and other CGIAR stakeholders. The Panel Report presents an analysis of the current activities within CGIAR concerning data, metrics and indicators, and offers a series of recommendations to address the key issues and challenges identified.

The study focused on the science underlying monitoring systems and was conducted in four steps:

1. An inventory of metrics and indicators in the CRPs and major partner initiatives within and outside CGIAR;
2. Analysis of gaps in metrics and indicators for monitoring CRP- and system-level IDOs;
3. Analysis of the comparative advantage of CGIAR to fill these gaps in relation to other research and development organizations;
4. Identification of opportunities to strengthen the science that underpins relevant metrics and indicator systems for CGIAR.

ISPC commentary on key findings and recommendations of the study

The Panel Report covers a broad range of issues relating to natural resource metrics. This is a rapidly developing field and many innovations are occurring that are of relevance to the New CGIAR. The study provides a wealth of information and detailed analyses on data, metrics and indicators as they relate to the CGIAR system, but it is inevitably incomplete as the range of activities in this field is so great. Numerous initiatives, both within CGIAR and in the broader community, evolved rapidly during the course of the study – including but not limited to CGIAR discussion of IDOs and the Open Access Policy and its implementation, and international work on the Sustainable Development Goals (SDGs). The goalposts were moving rapidly. The ISPC appreciates the study and the importance and relevance of the topics discussed. The ISPC aims to highlight some of the most important conclusions of the study and their future implications.

While the Panel Report advocates the development of sets of metrics that are 'simple and robust,' the panel had neither the mandate nor the qualifications to design a metrics and indicator system for CGIAR – that must be the task of the CRPs and the CGIAR Consortium. The study focuses, rather, on the issues that have to be considered in identifying and applying appropriate indicators, and commenting on particular indicators – such as time recording, crop yield gap, sustainability, full-chain nitrogen-use efficiency – that are either being used or have been proposed. The panel also places emphasis on the need for simple, direct measures and exposes the potential weaknesses of some composite indicators (discussed in relation to the example of total factor productivity, TFP). The panel concludes that indicators must be simple and robust, selecting the ones that "involve the fewest (hidden) assumptions and that are easy to understand and communicate to a wide audience." The panel draws attention to the fact that many indicators suggested for the SDGs do not meet these criteria. At the same time, the panel recommends that "the data, metrics and indicators procedures within CGIAR be comprehensive and address all dimensions of research and development activity," and that "the needs of both a learner focus and accountability are fully recognized in the data, metrics and indicator system," and that the "CGIAR data management system needs to be sufficiently flexible," and "handle data and metrics at various spatial (from the farm to the planet) and temporal (short- to long-term metrics) scales." Reconciling these diverse goals and dimensions remains the key challenge for CGIAR.

The Panel Report spans four separate sets of issues, which are discussed successively in the following sections.

A. Conceptual review of evolving practices in data, metrics and monitoring

This section provides information and analysis on a series of key issues.

- *Global and national assessments and monitoring.* The Report discusses the possible role of CGIAR in monitoring the state of the world's agricultural and natural resources systems. The panel recommends that CGIAR can contribute to the improvement of monitoring systems, and should put more effort into influencing global monitoring initiatives but should not take responsibility for these global initiatives. **The ISPC agrees with the conclusion that CGIAR should focus its own metrics systems on its own requirements for learning and accountability. CGIAR has an additional role in contributing concepts and data to those organizations that have a mandate for maintaining global data sets, such as the Food and Agriculture Organization of the United Nations (FAO) and other UN agencies, the World Bank, and the Demographic and Health Survey (DHS). It should also work with its national partners, guiding them in their work compiling national-level data sources. CGIAR should not duplicate the work of these other agencies.** The focus of CGIAR itself should be to develop monitoring systems that assess research impact and also increase understanding of trends, problems and opportunities that can lead to greater understanding of the dynamics of agriculture and help focus the work of CGIAR.

- *Data – changing needs and expectations.* The panel introduces a very useful discussion of 'big data' and emerging data streams, and suggests that CRPs and CGIAR Centers could contribute to this movement by making available and conducting analyses that yield more parsimonious and accurate metrics/indicators related to productivity, food security and dietary intake, among others. Of particular importance is **the point made by the panel on the possibilities of combining a small set of relatively standardized key SLO indicators with more real-time data on key drivers of change and/or intermediate results outcomes. ISPC endorses the suggestion that CGIAR should take advantage of the potential that information and communications technology (ICT) tools and systems thinking offer to create a learning system driven by evidence and used by a wide and diverse stakeholder body.**
- *Access to research data.* The discussion on this subject is particularly important as it articulates opportunities for linking open access initiatives with metrics and monitoring. For instance, the HarvestChoice project provides an excellent example of a landscape-scale evaluation framework for organizing key agricultural data layers into a standardized matrix across Sub-Saharan Africa. This platform, which compiles various biophysical and socioeconomic data sets and allows visualization and analysis of the mix of farming, cultural and socioeconomic conditions, offers an important resource to CRPs for priority-setting and for learning. These project-based data products should be expanded and integrated to serve as building blocks for implementing the CGIAR Open Access and Data Management Policy (CGIAR Consortium, 2013a), which has been commented on separately by the ISPC in July 2013 (ISPC, 2013d).[1]
- *Indicators should be "selected with a clear* ***sense of the theory of change****."* This point, which is made often in the Report, is extremely important as it also enforces the analysis and recommendations included in the ISPC white paper on theories of change and impact pathways (ISPC, 2012c).
- ***Data visualization techniques*** *and their power for articulating/making apparent relationships for research and policy-makers.* The short discussion on this topic is of great relevance for CGIAR. Although the CRPs may already be taking advantage of these techniques, this may still deserve attention from the CGIAR Consortium as a means to communicate (internally and externally) more effectively on issues of system-wide importance. The example given in the Report of the use of 'amoeba' diagrams is just one of numerous options for attractive user-interfaces for CGIAR metrics.

1. The ISPC also commented on a Consortium Office proposal to the Fund Council of March 2014 entitled *Supporting Open Access & Data Management Implementation* in April 2014, and then on a revised version of the proposal in July 2014.

B. Review of CGIAR plans for outcome and impact monitoring

This section reinforces the conclusions of the recent ISPC panel on 'open access and data management'. The ISPC notes the progress being made in this area by the Consortium and that it is in general consistent with the Panel Report.

The discussion of candidate metrics for SLOs highlights the high degree of uncertainty in this process and the difficulties that will be encountered in developing metrics for the SLOs. This is an area where greater realism and pragmatism will be required. Simple metrics to track CGIAR contributions to the achievement of the very broad and ambitious SLOs will remain a major challenge. The ISPC concurs that a diversity of indicators will be required for the SLOs; tracking CRP performance and contributions closer to the research activities will be needed, but to demonstrate the relationship between the research and development impacts emphasis must also be given to identifying the associated but higher-level indicators which can illustrate the dynamics of the impact pathways and test theories of change.

C. Alignment with the SDGs

The panel highlights the desirability but difficulty of aligning the CGIAR metrics system with the SDGs. The SDG process remains very fluid and it is still unclear what the UN System will finally adopt in 2015. At present there are parallel processes developing different sets of SDGs. Most observers predict that the official SDGs will be set in terms that are too general and all-embracing to be of much direct relevance to CGIAR. The Panel Report focuses on the Sustainable Development Solutions Network (SDSN) papers, which are currently out for review, and provides useful insights into the relevance of these for CGIAR. However, the final SDGs may be quite different and will likely set very general aspirational targets. The ISPC sees value for CGIAR in understanding and describing how the SLOs may align with the global SDGs and targets, but does not believe that the SDGs will provide a basis for monitoring.

D. Conclusions and Recommendations

The Panel Report acknowledges CGIAR's overall contribution to data and metrics on agricultural systems and highlights the lack of system-wide capacity for collecting, archiving and storing data. Hence the panel recommends that CRPs should have adequate provision for curation, quality control and archiving of data and for making data, metrics and indicators available to partners and users. **The ISPC endorses the recommendation that the Consortium provide a normative and control function and ensure periodic peer review of data, metrics and indicators. The provision of comprehensive easily accessible high-quality data and metrics on agricultural systems should be one of the major public goods products of CGIAR.** However, the ISPC in its commentary on the Consortium's draft Open Access Policy (ISPC, 2013d) **advised caution against establishing a centralized system 'to oversee data, metrics and indicators throughout the system.' The Consortium has to ensure that the system of metrics is maintained and that it has a role in peer review and quality control, but the ISPC and the panel both favor a decentralized network of data hubs and not an overly centralized system.** Semantic webs and controlled vocabulary (ontologies) can enable a large degree of decentralization in information networks.

The panel strongly approves of the intentions behind the new **Open Access and Data Management Policy** developed by the CGIAR Consortium (CGIAR Consortium, 2013a). The panel recommends that the main structure of the open access arrangements be put in place rapidly, but given that change is so rapid in this domain, it will be important to build in flexibility and the ability to adapt to changing circumstances as the policy implementation proceeds. The ISPC supports this recommendation and notes that guidelines have been issued by the Consortium and cross-Center/CRP teams are already moving forward with implementation of the policy.

To build a CGIAR resource for the future, the panel thinks it is essential to measure the basics, and to measure them well. The portfolio of CRP activities has expanded dramatically in recent years, and research is being conducted by very large consortia composed of diverse types of institutions and users. A focus on 'systems' that has been adopted within a number of CRPs creates its own special needs and opportunities for data, metrics and indicators. Where more complex metrics and indicators are presented the panel proposes two principles: (i) All input data should be standardized and made available, including provision of details of the methods used for data collection; and (ii) The calculations used to derive metrics and indicators should be presented transparently. While ISPC supports the panel's call for transparency, it is not convinced by the call for "a comprehensive open CGIAR database" or that "all input data should be standardized." There is an apparent tension between the call for a comprehensive approach and the need for simplicity. Clearly, data required for necessary metrics and indicators at the CRP and system levels need some minimum level of standardization. Thus (as intimated above) in contrast to the 'comprehensive approach' to the six main purposes listed by the panel (see Executive Summary, page 10), the **ISPC suggests a different (possibly more pragmatic) strategy that would arise from considering how best to achieve each of the six purposes** for metrics and indicators. Such a disaggregated (rather than comprehensive) approach is in line with the ISPC's commentary on the Open Access Policy, as CGIAR moves forward with data, metrics and monitoring. In this same vein, while it might be subsumed under purpose (i) 'research tool,' or purpose (vi) 'foresight and exploration,' the ISPC considers that identification of extrapolation domains and data to address issues of external validity might have been included among the purposes listed.

The panel makes a distinction between different metrics required for learning and for accountability throughout CGIAR. While metrics for accountability will be linked to predetermined outcome targets, a learning focus will require the tracking and detection of patterns in metrics accumulated over time. **The ISPC endorses the panel's recommendation that the needs for both a learning focus and accountability are recognized in a data, metrics and indicators system, and that no single objective should dominate.** As stated above, a pragmatic purpose-by-purpose approach should be considered for establishing priorities.

CGIAR needs to handle data and metrics at various temporal scales and at spatial scales ranging from the farm to the globe. The panel recognizes the scientific challenges associated with the scale issues and metrics aggregation/disaggregation. This issue was also identified by most CRPs as one of the major challenges they are facing in developing metrics and indicators (see the results of the survey questionnaire, Annex 2), and was discussed during the ISPC metrics workshop in December 2013 (ISPC, 2013e). **The panel recommends special attention be given to the problems of aggregation and disaggregation of data collected at different spatial and temporal scales.** Indicators to be proposed by CGIAR should specify the scale at which they are relevant and indicator use should be assessed empirically for validity. The panel draws comparisons with a multi-level analysis system, known as SEAMLESS, that was designed for policy decision support in the European Union. A similar approach for developing a multi-level system could allow better integration of data and metrics across CRPs to allow analysis of trade-offs and interactions across system components, but such systems are costly to establish and maintain.

The panel recommends that the CGIAR Consortium develops a comprehensive ontology for data and metrics systems within CGIAR, with contributions from all the CRPs. The ontology should not be developed in isolation from data and metrics work being conducted by other organizations. This is in agreement with the intent of the most recent proposal by the Consortium for the introduction of open access and data management, including the use of controlled vocabularies, and is supported by the ISPC.

The panel points out that the development of a system of data, metrics and indicators is occurring at the same time that the **Strategy and Results Framework (SRF)** is being refined and that the links between this and the SLOs and IDOs are being made explicit. The SRF management update (CGIAR Consortium, 2013b) suggests that the IDOs should be achievable in 9–12 years, a shorter time frame than the 15–20 years for the SLOs. In some cases there may be direct linkages and, for others, the achievement of SLOs could be an aggregated outcome of the IDOs. The panel considers it important that work on data, metrics and indicators is well coordinated with the process of developing the SRF and the outcome targets. **The ISPC supports the clarification of the principles on data, metrics and relationships between indicators in the new SRF.**

The panel **recommends pragmatism in the use of metrics to measure progress toward these outcomes (IDOs and SLOs). At present, the metrics debate may be excessively focused on SLOs and IDOs.** Accountability metrics should focus in the short term on immediate development outcomes and recognize the difficulty of addressing the needs of measuring long-term SLOs. The ISPC agrees with this conclusion and suggests that every CRP should selectively invest in efforts to build credible trials for (*ex post*) impact assessment where successful interventions are tested.

Further, in relation to impact assessment within CGIAR, the panel welcomes the new Special Program for Strengthening Impact Assessment in CGIAR (SIAC), which recognizes the need for broadening the range of metrics used to assess impact. In particular the panel thinks that extra attention should be paid in future to the measurement of impact on natural resource systems. **The panel endorses the work of the Standing Panel on Impact Assessment (SPIA) and encourages it to prioritize the issue of evaluating natural resources management (NRM) research projects.** The ISPC agrees in principle with this recommendation, encourages the work of groups convened by the Consortium to come to a resolution on early indicators and targets, but notes that addressing all difficult measures (such as educational measures of empowerment) requires specific and focused research work by relevant (and perhaps specifically mandated) groups across CGIAR.

The panel notes that a key use of the impact evaluation system is to **prioritize research investment**. While the Panel Report is, however, relatively silent on adequate measures that might be adopted at the level of development outcomes, it does raise one of the dilemmas in the use of indicators in priority-setting. When agricultural technologies (e.g. new crop varieties) are adopted, the poorest often benefit least in absolute terms, though they may benefit most in relative terms – as a proportion of their income or an increase in months of food self-sufficiency. **The panel recommends that CGIAR should not retreat from working on difficult problems based on arguments couched in simple economic returns**. The ISPC considers this as an example of the case noted earlier, that direct economic (and other) benefits from the results of CGIAR research should be measured by programs as well as defining indicators (at the CRP domain or system level) to monitor the relationships between local and national or regional development (see also the discussion of big data and drivers, in the second point in Section A of this commentary).

The panel notes that substantial emphasis is rightly placed on a wide range of **partnerships** through which CGIAR will achieve its goals, which also means that CGIAR is dependent on the performance of both research and development partners in achieving impact. The panel questions whether the new prioritization around impact would mean that CGIAR will avoid weaker partners, such as the national agricultural research and extension systems, that often suffer from chronic underfunding. Yet building the capacity of such partners could be critical for the long-term sustainability of research outcomes. The ISPC encourages the Consortium to advance a strategic discussion of partnership arrangements and capacity building in support of impact delivery as a key component of the SRF.

The development of the CGIAR system of data, metrics and indicators is occurring at the same time as numerous other initiatives with similar or overlapping objectives. The panel urges that it **is important that CGIAR takes note of these initiatives and, to the extent that it is appropriate, aligns its own work with them**. The ISPC concurs and strongly supports the need for the setting of research performance indicator targets for CRPs and their linkage to higher-level development outcomes using consistent vocabulary, to maximize the comparability and relevance of the system's efforts. This will require continuing coordination of efforts in this domain from the Consortium and CRPs over and above the effort on Open Access that they have recently embarked upon.

Revision of 27 October 2014

Independent Science and Partnership Council

II.

Panel Report:

Data, metrics and monitoring in CGIAR – A strategic study

Preface

The Independent Science and Partnership Council (ISPC), as part of its mandated work on strategy and trends, commissioned this strategic study on data, metrics and monitoring in CGIAR in mid-2013, during a time of intense CGIAR activity on data, metrics and indicators. All 15 of the CGIAR Research Programs (CRPs) are developing metrics and indicators. The CGIAR Consortium Office is revising the Strategy and Results Framework (SRF). Simply reading and assimilating all of the numerous detailed reports has been a humbling task for the panel of experts who undertook the study. Hundreds of very capable CGIAR scientists are engaged in this work, and the panel recognizes the danger of simply adding another layer of analysis to what is clearly a rapidly moving target. Weaknesses in metrics and long-term data management in CGIAR have elicited concern for several years. We (the study panel) hope that our reflections can highlight priorities and pitfalls in developing a robust and coherent approach to data, metrics and indicators across the diverse realms of research for development within CGIAR.

The study was undertaken by a panel of external experts, chaired by Dr Ken Giller (Professor of Plant Production Systems, Wageningen University, the Netherlands) and composed of Dr Simon Bell (Professor of Methodology and Innovation, Open University, UK), Dr Nancy Mock (Professor at Tulane University, USA) and Dr Robert Hijmans (Professor at the University of California, Davis, USA). The panel worked under the guidance of ISPC member Dr Jeffrey Sayer (Professor in the School of Earth and Environmental Sciences, James Cook University, Cairns, Australia) and with the assistance of Dr Rachid Serraj from the ISPC Secretariat. The study was designed to address the terms of reference found in an ISPC concept note (ISPC, 2013a). The study included a survey on the planned use of metrics and indicators by the CRPs, including CRP- and system-level outcomes, and a workshop with CRP members, selected partners and donors, and CGIAR stakeholders.

The Study Panel

Executive summary

Observation, measurement and the detection of patterns in data are the basis of scientific research. Advances in methods of measurement and analysis often lead to new scientific understanding. A key global research organization such as CGIAR has a leading role to play in assembling, managing, analyzing and transmitting data, metrics and indicators concerned with agricultural research and development. Since its beginnings, CGIAR has collected data on agricultural and natural resource systems. CGIAR has some notable achievements in establishing and maintaining databases and analytical models and has contributed to global data sets managed by others. But there is no standardized approach to data management and there are problems of data quality, storage and retrieval. Much of the data collected in the past can no longer be retrieved or used, suggesting that resources of immense potential value for understanding agricultural development have been lost. This problem was already well articulated in the CGIAR Science Council's earlier Stripe Reviews of Social Science and Natural Resources Management research in CGIAR (see Annex 8).

Donors and other stakeholders of CGIAR expect clear evidence of important results from the increased investments following the recent reform process. In particular, CGIAR needs to show that its work is having a positive effect on the system-level outcomes (SLOs) of reducing poverty, increasing food security, improving health and nutrition, and the sustainable management of natural resources. To do so requires that CGIAR should quickly establish a system that documents the outputs, outcomes and impact of the system. The study panel recognizes the need for speed but does not underestimate the cost and complexity of building such a system and we thus advocate a gradual approach. However, this report is not restricted to metrics for the assessment of accountability and impact. The emerging Strengthening Impact Assessment in CGIAR (SIAC) program, led by the Standing Panel on Impact Assessment (SPIA), will focus specifically on impact metrics (CGIAR, 2014). The present report casts the net more widely and looks broadly at the issues of metrics and indicators in the areas of work of CGIAR and in agriculture, rural development and natural resources management (NRM) in general.

The recent adoption of a CGIAR-wide, Open Access and Data Management Policy provides a unique opportunity for the development of improved data management systems and many of the elements for such systems have been described in the policy document (CGIAR Consortium, 2013a). Some of these aspects are elaborated later in this report.

The study panel's basic thesis is that CGIAR could greatly benefit from a system for collecting, curating and archiving data and metrics as a resource for learning, research, policy-making, priority-setting and impact assessment. A second principle is that CGIAR should align with, contribute to and exploit opportunities provided by numerous other initiatives occurring outside of CGIAR; these opportunities are explored in detail in later parts of the report.

The report recognizes the following key issues and offers guiding principles and recommendations for addressing them.

1. *The purpose of metrics.* Recent demands for improved metrics have been mainly driven by the need to have better measures of the impact of CGIAR research on the SLOs. This report gives special attention to 'accountability' metrics and indicators but notes that there is a danger that the current emphasis on short-term impact metrics might divert attention from the need for more comprehensive, long-term data management systems to meet the needs of research, monitoring and impact assessment in an integrated way. A balanced system of common and compatible metrics is needed to serve six main

purposes: (i) as a research tool to understand the dynamics of change around agriculture, food systems and natural resources; (ii) to provide comparability that allows reporting in an aggregated and CRP-specific manner across CGIAR; (iii) to predict and measure research impact on the SLOs; (iv) to assess causal linkages and trade-offs among research impacts on SLOs; (v) to learn from failures and successes to allow for the development of more effective research projects; and (vi) to provide the opportunity for foresight and the exploration of future trends.

2. *Use of metrics and indicators.* In designing the metrics and indicators system, emphasis should be placed on learning – by researchers and donors alike – about the dynamics of agricultural systems and the return on investments in research. A robust metrics and indicators system should be at the core of CGIAR research and planning. If well designed, a learning-based system can allow reporting as part of a single framework. The focus on four SLOs allows for the development of a relatively simple system. However, metrics must not become a straitjacket; they should embrace the diversity and complexity of CGIAR programs and should not restrict them.
3. *Data management.* Although there are several positive exceptions, CGIAR has a weak overall record on long-term data management. Data from some past research efforts are difficult or impossible to access, and in some cases the data are of poor quality or inadequately described and curated. Data sources are distributed throughout the CGIAR Centers and CRPs and there are no uniform standards or archiving protocols. Ground rules that apply across the range of CGIAR activities have been established in the Open Access and Data Management Policy. A degree of coordination in data management, quality control and curation, for example to develop shared ontologies and standard templates for data collection, will be essential to making the system reach its potential and to meet the needs of users. These tenets have since been embodied in the Consortium's proposal for the implementation of open access and data management in CGIAR. The panel favors a 'distributed network' of data hubs linked by minimal superstructure. Additional contractual requirements may be required to ensure that data are made available in a timely manner.
4. *Impact assessment.* CGIAR should have a simple and robust impact assessment system. Major investments need to be supported by ***ex ante*** analysis. Such assessments can be updated as research progresses (or not) and as data and methods improve. *Ex post* impact assessments should also be carried out more regularly, including for less successful projects. It is important that these studies are coherent and transparent. All basic information should be made available on a website that gives access to the raw data and methods used to compute metrics, indicators and impact.
5. *Global and national assessments and monitoring.* CGIAR needs to clearly define its role in monitoring the state of the world's agricultural and natural resources systems. There are at present several initiatives aimed at improving public, national and international monitoring of agriculture, health, poverty and natural resources. The need for increased monitoring is recognized in the current attempt to define internationally agreed Sustainable Development Goals (SDGs). Standardized international data sets are costly to establish and maintain. In most cases, CGIAR cannot and should not take responsibility for them. However, CGIAR can contribute to improving monitoring systems, and should put more effort into influencing global monitoring initiatives to ensure that its own needs are met. CGIAR should work with organizations like the Food and Agriculture Organization of the United Nations (FAO) and other UN agencies, the World Bank and the Demographic and Health Survey (DHS). In particular, CGIAR should participate in compiling national-level data sources. It can also conduct monitoring under specific circumstances where it has a comparative advantage to do so.
6. *Alignment with other metrics initiatives.* CGIAR should participate in the development of the indicator system that is being established to monitor progress on the SDGs. Although the alignment of the CRPs

with the SDGs is an attractive notion, the approach being taken to select indicators for the SDGs will result in indicators that are more generalized than those required by CGIAR.

7. *Focus resources on establishing basic key metrics.* Given the emergence of 'data science,' CGIAR can play a leading role internationally, particularly in developing countries. Building on open access data, CGIAR can use new analytical techniques that are available to support interdisciplinary research. To create a resource for future research, the panel recommends a focus on measuring the fundamental attributes of agricultural systems in a robust and transparent manner (e.g. crop yields, livestock numbers, farm size, and household composition). Such basic metrics supply the underlying data for calculating metrics and indicators at higher levels. The methods used for collecting such information and the assumptions made must be clearly stated. This will ensure that the data are 'time-proofed' for revisiting and recalculating metrics and indicators in the future.
8. *Rationalize the investment in baseline surveys.* The information collected from the thousands of studies and surveys conducted in developing countries over the past decades could provide a rich picture of the dynamics of agricultural systems. Where CGIAR invests in surveys to establish baseline data or for other purposes, a more systematic effort should be made to allow the compilation of survey results over space and time. The CRPs should also fully exploit opportunities for economies of scale in collecting and sharing data among themselves.
9. *Key scientific issues needing further research. These include:*

   (a) how to 'aggregate/disaggregate' metrics from the project to program to system level. Scaling metrics and indicators is a branch of science in its own right. Far greater attention needs to be given to linking metrics and indicators across spatial and temporal scales, as well as across different levels along an impact pathway. This requires the work to be embedded in a robust theory of change and for careful consideration of the methods used for aggregation;

   (b) development of metrics for all major fields of CGIAR activity. Until now, metrics are mostly lacking for capacity building (in its broadest sense), for certain aspects of NRM, for innovation systems, social learning, empowerment and the capacity to innovate. While some of these fields may not lend themselves to simple or routine monitoring, CGIAR needs to evaluate where new metrics could be useful and when they might not be;

   (c) trade-offs and interactions;

   (d) the cost-effectiveness of indicators.

1. Introduction

To meet its goal of a food-secure world, CGIAR needs to understand past and future trends in agriculture, rural development, health, poverty and natural resources use, and the ways in which these trends are affected by research. Such knowledge underpins the research process. Knowledge of status and trends in systems can be used to improve research, guide research investment and make CGIAR more efficient and effective. Good systems for managing data, metrics and indicators are thus critical to the overall success of CGIAR. Open access repositories of data, metrics and indicators of agricultural systems should be a major product of CGIAR research and a major resource for learning and future research.

Since the recent reorganization of CGIAR and the consolidation of research into CGIAR Research Programs (CRPs), there has been an increased demand for evidence that these CRPs are effective. After its review of the 15 CRPs, the ISPC questioned how the contribution of CGIAR to the goals of the Strategy and Results Framework (SRF) would be assessed and how progress toward those goals would be monitored and tracked. The CGIAR Consortium and its funders have recommended the identification of metrics to measure CGIAR's success in implementing the SRF and to connect the performance of the CRPs to the higher goals of the system-level outcomes (SLOs).

The renewed interest in metrics within CGIAR is occurring at a time when the world of data collection and analysis is undergoing rapid changes. New methods and opportunities are emerging that could support CGIAR research. Emerging monitoring systems include methods not only to assess research impact but also to increase our understanding of trends, challenges and opportunities – this could help to focus the work of CGIAR.

Although many stakeholders appear to agree on the need for increased measurement, debate continues on what should be measured, how, where and by whom. Key questions that CGIAR needs to address include the following.

- At what level should CGIAR engage in data collection and monitoring?
- For what purpose?
- What can be done by others?

The goal of this report is to provide some guidance in answering these questions.

The study was undertaken by a panel of external experts. The review described in these pages focused on the science underlying monitoring systems and was conducted in four steps:

1. An inventory of metrics and indicators used in the CRPs and major partner initiatives both inside and outside CGIAR;
2. Analysis of gaps in metrics and indicators for monitoring CRP- and system-level outcomes;
3. Identification of the comparative advantage of CGIAR to fill these gaps in relation to other research organizations;
4. Identification of opportunities to strengthen the science that underpins relevant metrics and indicator systems for CGIAR.

1.1 What do metrics mean in the context of this report?

In this report we discuss the role of data, metrics and indicators in CGIAR. Data occupy a loosely defined hierarchy, from least-transformed (observations, raw data) to most-transformed (estimates, indicators). Indicators should be informative to support decision-making. For this reason, data can be an indicator in one context but could be considered a metric or even raw data in another.

- Raw data are observations, such as weight, height, plot size.
- Metrics are computed by aggregating and combining raw data, for example, yield or height-for-age. They often represent the values on which indicators are built.
- Indicators are summary measures that reflect system properties. Examples include infant mortality rates and the prevalence of acute malnutrition, or changes in these values.

There are no standard definitions of data, metrics and indicators in the literature despite the fact that they are the building blocks of any results management framework. A results management framework describes the goals and intended outcomes of projects and investments. These are then translated into indicators of outcomes, often termed 'results,' which are built by collecting data, calculating metrics and then combining/reducing these into indicators. Essentially, a metric becomes an indicator when it is used for decision-making; thus all indicators are metrics, but not all metrics are indicators.

See Annex 1 for definitions of relevant terms used in this report.

1.2 Why are metrics needed?

A balanced system of common and compatible metrics serves six main purposes: (i) to provide a tool for understanding the dynamics of change around agriculture, food systems and natural resources; (ii) to enable comparability that allows reporting in a CRP-specific and aggregated manner across CGIAR; (iii) to predict and measure research impact on the SLOs; (iv) to assess causal links and trade-offs among research impacts on SLOs; (v) to enable learning from failure and success thus ensuring more effective research projects in future; and (vi) to enable prediction and exploration of future trends.

The recent debate on metrics in CGIAR has been characterized, on the one hand, by discussions of indicators for monitoring CRP progress (e.g. at benchmark or sentinel sites) and for measuring the short-term 'impacts' of CRPs through periodic evaluations, and, on the other, by discussions about how to frame the indicators of the higher-level development outcomes that the CRPs and the system as a whole seek to achieve. The adoption of new terms for the higher-level outcomes (i.e. intermediate development outcomes, or IDOs; and SLOs), which were introduced to increase program focus, may have distracted from an overarching and coherent approach to indicator development in CGIAR.

2. An assessment of the current issues for CGIAR in developing a coherent framework

In this first part of the report, the panel reviews some of the current issues arising from CRP experience, which were mentioned in a survey questionnaire and discussions at the metrics workshop in December 2013 (ISPC, 2013e). **The panel's review of documents and the survey of CRP managers reveal a lack of coordination across the CRPs in terms of baseline data collection and site characterization. There is inconsistency in the definition of geographic domains, criteria and benchmarking across the CRPs.**

2.1 A summary of responses to a survey of CRPs on data, metrics and indicators[2]

The CRPs are addressing the issues of metrics and indicators for research in diverse ways, yet CRP target domains overlap. Most CRPs have engaged in site characterization and baseline data collection. For example:

- The CRP on Climate Change, Agriculture and Food Systems (CCAFS), one of the first CRPs to be launched, selected sites in three of its five targeted regions. CCAFS has defined indicators and collected baseline data in all sites. Baseline data collection for the other two regions will be completed by the end of 2014. CCAFS has expressed its willingness to modify its approach to collecting data if CRP-wide agreements can be reached.
- The Forests, Trees and Agroforestry (FTA) CRP has set up a network of nine sentinel landscapes in which a core set of metrics is being measured.
- The Water, Land and Ecosystems (WLE) CRP has not yet started collecting baseline data for various reasons, including the need to adopt the metrics of their – still to be identified – boundary partners rather than collecting data relevant only to the CRP.

The survey indicates that individual CRPs are using a number of methods for collecting data, including household surveys, remote sensing, on-station and on-farm trials, as well as harvesting existing databases. Some CRPs (e.g. Livestock & Fish) are collecting baseline data within the context of randomized controlled trials (RCTs) with the aim of being able to attribute the contributions of their respective programs to achieving the desired outcomes by comparing data from program locations with control locations. Depending on their targets, some CRPs are relying on data sets from earlier projects and collecting specific metrics on areas such as productivity, natural resources management (NRM), nutrition and gender. Some CRPs (e.g. CCAFS) have designed metrics to track higher-level outcomes such as food security, livelihood status, adaptation and mitigation actions, and emissions.

The CRPs are variously relying on secondary data sources available in the public domain (e.g. statistical offices, FAOStat, national data sources). Most are using the World Bank Living Standards Measurement Studies (LSMS), although these only exist for some countries. The CRPs have also established new partnerships for their specific data, metrics and indicators needs. For example, Livestock & Fish (L&F) and Agriculture for Nutrition and Health (A4NH) are working with IFPRI and the World Health Organization (WHO) on indicators of nutrition and health, such as dietary diversity indices (Annex 4). A few scientists

2. See Annexes 2 and 3 for a summary of CRPs responses to the survey questionnaire.

are participating in the development of the Sustainable Development Goals (SDGs) and related indicators, although the role of CGIAR in this process is limited.

There are continuing efforts to enhance coordination among clusters of CRPs (e.g. the systems CRPs, the CRPs working on natural resources, those working on selected commodities). Examples of joint activities initiated by clusters of CRPs include:

- The systems CRPs are developing common research plans in areas of overlap, e.g. targets and indicators for 'capacity to innovate,' as a common IDO.
- The three NRM CRPs are working jointly on developing metrics, indicators and frameworks for the IDO on 'adaptive capacity.' However, they face a challenge in identifying the scope of these indicators and the conditions under which the framework should be applied, given that the cost of collecting the indicators is likely to be prohibitive.
- There is a significant effort to coordinate CRP activities in Burkina Faso, where a common monitoring plan is being developed by WLE, FTA and CCAFS, together with some of the commodity CRPs, such as Grain Legumes and Dryland Cereals.
- A joint study on 'Gender Norms and Agency in Agriculture' is being carried out across all of the CRPs with the aim of harmonizing gender indicators through the CGIAR Gender Network.

These efforts will need to be more systematic to enhance coherence and synergy across the CGIAR research portfolio and target domains.

2.2 How the CRPs have approached program-level indicators

The review revealed that progress in developing CRP indicators and metrics for monitoring progress across levels and outcome categories is generally at an early stage (as of October 2013; see Annex 3). For tracking productivity outcomes, most CRPs suggest using yield and adoption indicators at the plot or household level, either as direct measurements of yield and number of adopters or expressed as percentage of yield and profitability increases. A4NH proposes that yield and adoption data be collected at field, plot, household or individual level (as in gender disaggregated data), but notes that it could be aggregated at higher scales. L&F will measure a series of specific productivity indicators, including annual milk or fish yield, meat yield per animal, annual kidding percentage, litter size, weaning percentage and animal mortality rate. The program plans to monitor the adoption of new or improved technologies and management practices across scales from the household to regional level.

At the landscape/district level, WLE plans to measure a series of indicators, including internal rate of return on investments (flood harvesting, groundwater management, new irrigation schemes, resource reuse and recycling techniques, etc.). FTA suggests measuring net primary production and land use at the field, village and landscape levels, whereas CCAFS proposes to measure a household productivity index and the number of changes in practices made at the field or household level, as well as monitoring community perceptions of changes in natural resources at the village level.

For tracking livelihoods, several measurements of welfare, income, and food and nutrition security have been proposed by CRPs to provide indicators across scales. These indicators include household income and intra-household food and nutrition security, percentage income increase, percentage increase in consumption, quantity and quality of target commodities supplied, dietary diversity, consumption of target commodities by the target population, level of awareness and attitudes toward dietary diversity practices. Similarly, a series of metrics and indicators is being envisaged for monitoring progress in social outcomes (e.g. empowerment, gender equality), environment and natural resources outcomes (e.g. soil health, water, biodiversity, climate-related greenhouse gas emissions) and policy outcomes (see Annex 3 for details and examples).

The A4NH team developed a concept note that illustrates how impact evaluation and projections about the scale-out of interventions might lead to goal-setting *ex ante* for SLOs and IDOs. The potential size of the effects resulting from A4NH interventions was estimated based on rigorous impact evaluation with counterfactuals. The scale-up factor was based more on less objective criteria. However, this basic type of analysis, where impact effect estimation is combined with a projection of the scale-up process, provides a methodology for estimating the IDO and even SLO contributions of CGIAR research. (See the table in Annex 4, which focuses on concrete nutritional outcomes and examples.) This type of analysis can be helpful for setting targets and evaluating progress toward meeting these targets. CGIAR also works with modelers, such as Hubbard Consulting, to develop forecasts of varying outcomes based on a range of differing assumptions about initial conditions and scale. These types of techniques, combined with rigorous impact evaluation to estimate effect size, can offer potentially useful tools to CGIAR for targeting its contributions to SLO and system-level IDO change.

The panel observes that **there is a great variety of approaches, metrics and indicators used by CRPs across levels and outcome categories in their mandated areas. In most cases, the work takes place in very small geographic areas. This may make it difficult to use data to track impacts at the country level. More thought needs to be given to understanding and measuring the links between small-scale measures and national-level changes in indicators of the SLOs.**

2.3 Gaps and challenges

The survey of CRP managers identified some of the major gaps and challenges faced by the CRPs in developing metrics (Annex 2). The metrics workshop also highlighted issues of concern to participants and areas for further research and development (Box 2.1). The major gaps identified by survey respondents were the following.

- The system CRPs are on uncharted ground, needing time to develop metrics. The metrics need to cover trade-offs and interactions across components in the systems (e.g. total factor productivity). Social indicators need to include cultural variables.
- Specific metrics are needed on community and individual empowerment, capacity to innovate and capacity to adapt.
- Geographic coverage is limited and there are significant data gaps. Care will be needed to ensure the representativeness of data, e.g. for secondary data.
- Thematic coverage: studies/methods differ in terms of metrics covered, level of detail and reliability; there are difficulties linking metrics and studies/methods from different domains, for instance biophysical versus socioeconomic and gender indicators.
- A major challenge is to aggregate/disaggregate metrics from project to program level.
- **IDOs that are specific to the CRPs show little synergy with respect to using existing data sources and methods. There is a need for cross-partner platforms, shared data platforms, and guidelines for compatibility across CRPs. The system-level IDOs can be useful for addressing this problem.**
- For some IDOs, such as gender, research is still needed to determine which metrics are best for specific contexts.
- Some indicator approaches are clearly better value for money than others. Realism is needed on the cost-effectiveness of indicators.

- Given that the IDOs are still being developed, there will be a need to revisit the baseline metrics in light of the final IDOs.
- We need to identify boundary partners, understand their needs and work with them to define metrics and indicators.
- The attribution of impact will always be problematic. The focus should be placed on identifying direct contributions from CGIAR research.

In the next sections, the panel considers principles for selecting indicators (Section 3), with a view to encouraging the adoption of compatible indicators across the CGIAR system (Section 4).

Box 2.1 Metrics workshop – Summary of group discussions

A workshop was organized by the ISPC and the study panel (Rome, 10–11 December 2013) to discuss the initial findings of the study with CRP representatives and selected stakeholders and partners (for details, see the Workshop Report: ISPC, 2013e).

After a series of presentations by panel members and various CGIAR stakeholders and partners, two group discussion sessions took place. In the first session, four working groups discussed the main issues of data, metrics and indicators from the perspective of the SLOs. General observations included the following.

- As a knowledge-based organization, CGIAR should be concerned about the contribution of agriculture to poverty reduction.
- We need to shift the focus from doing research on farmers to doing research with farmers.
- There are subjective versus objective measures of poverty, but the main question at the research outcome level should be: how many people have adopted our technologies?
- There is an inherent value in the adoption of a technology.
- Two thresholds for adoption: (i) When are the potential adopters capable of taking risks? (ii) What do we mean by lifting people out of poverty – social and/or financial poverty?
- We need metrics at different levels within a more complex impact pathway and theory of change, with line of sight between project outputs, CRP results, IDOs and SLOs.
- The theory of change we adopt may need adjustment over time. It needs to capture indicators at different points along the impact pathway, and reflect both short- and long-term time frames. It needs to focus on impact pathways that are both operational and monitorable.
- Success at the level of research outputs depends on the scientific review process – but at the outcome level, success will depend on the clarity of impact pathways and the underlying assumptions. We need to build an understanding of these complexities, supported by research on the process of delivery, and to generate evidence on how the different levels feed into each other.
- We have agreed on IDOs but we do not have standardization of metrics; we need to understand where the responsibility for IDOs lies within the CRPs.

Building on the momentum of the first session, a second group discussion session focused on the following themes summarizing key challenges and gaps identified in the survey.

1. Trade-offs and interactions across systems components

- Trade-offs can be assessed when their impacts on all SLOs are estimated, so all of them should be part of a common impact pathway.
- Trade-offs will always happen, so we need to be aware of them; we should analyze whether the cost of the trade-off is worth it, and according to whom.

- Impact pathways should not be developed by researchers alone; donors and other stakeholders should be involved in the process.
- Specific theories of change should be developed for each research theme or major research activity. These theories of change are agreed impact projection models that are useful for planning and can be used for evaluation later.

2. Metrics for empowerment and capacity to innovate

- Both types of metrics are hard to quantify. Do people make their own choices? Are people/societies able to do/adopt new things?
- Most credible and easy-to-measure indicators seem quite unable to answer these questions; there is no easy way to define indicators on empowerment and capacity to innovate at this stage.
- Measure the outcomes, e.g. in measuring capacity to innovate, one could perhaps examine how innovative national agricultural research systems (NARS) have become as a result of their relationships with CRPs.
- Capacity to innovate results from properties of both the intervention and the recipient population. So for an innovation, one could estimate the requirement for capacity to innovate.

3. Criteria for selection of indicators

3.1 Simple metrics

There is a strong tension between simple metrics and indicators that can be measured in a robust way, and indicators that give more detailed insights. A good example is the measurement of productivity in agriculture (see Box 3.1).

Box 3.1 Measuring productivity

Measures of productivity are important indicators of agricultural performance and have been selected as one of the common IDOs. Productivity can be defined as yield per unit area, yield per unit labor or yield per unit of another input (nutrients, water), although the most common metric is yield per unit area. The measurement of productivity is often based on farmers' estimates and recall from past seasons. Apart from the lack of record-keeping by farmers, which can lead to difficulty in accurate recall, many factors may confound an accurate measurement of productivity.

Farmers' estimates of production per field are often reported in local units, such as number of bowls, buckets, 50 kg sacks or wheelbarrows. The calibration of such measurements is essential as the container may not be a standard size and may not be filled evenly. Crop yield can be expressed as harvested crop (before threshing/shelling/dehusking), as cleaned grain or as economic yield. Yields can be expressed as dry weight or at 14 percent moisture. If not measured consistently, when combined such factors can easily result in major errors in the estimation of yields.

Potential pitfalls in crop area

Many smallholder farmers do not have accurate knowledge of the area of the land they own and farm, particularly those who do not have formal title to their land. Cropped area can be defined in many ways: the area planted to a given crop in the previous season; the area of land that could have been cropped if sufficient labor/mechanization were available; or the area of land set aside for cropping, including land left fallow in the season in question.

Farmers' estimates of land area are often subject to rounding errors. Using GPS to measure field and farm areas produces more accurate results than relying on farmers' recall, although it is more time consuming (Carletto *et al.*, 2013a, 2013b). In addition, fields that lie far from the homestead (where interviews usually take place) tend to be off the GPS grid (Kilic *et al.*, 2013).

Many CGIAR Centers have considerable experience in conducting surveys and have surely encountered and addressed such issues in the past. Overall, farmers' recall alone cannot be relied upon for measuring productivity and need triangulation and checking using actual measurements. The key finding of Carletto *et al.* (2013b) is that the errors generated by relying on farmers' recall are systematic: farmers with smaller land holdings tend to overestimate their land areas and those with larger land holdings underestimate their areas, leading to bias in area–productivity relationships.

3.2 Composite indicators

Indicators can either be used directly or combined into a composite index. A composite index is often derived from a set of 10 or more metrics, which are given different relative weightings depending on their perceived importance. Composite indices are useful when dealing with systemic qualities (e.g. welfare, happiness, development or sustainability), when single indicators cannot adequately assess the complexity of the context. However useful, composite indices tend to mask underlying trends and the weighting systems may be arbitrary. The same value of the index can be derived from different combinations of values of the underlying metrics. This makes it difficult to understand the meaning of changes in the index.

The search for a systemic and multi-dimensional portrayal of a complex reality results in indicators being combined in composites and as such present arrays of linked but conceptually segregated domains. A composite indicator should, if it is effective, reveal the results of the array in a single indicator or event number ("the answer is 42!"). To work, the composite should be underpinned by a conceptual structure that allows different indicators to be included and weighted. The composite should represent the qualities and values of the item being studied.

Composite indices often cover a range of domains, for example the Women's Empowerment in Agriculture Index for measuring empowerment, agency and inclusion of women in the agricultural sector, and the Human Development Index, which includes indices on health, education and living standards. These two examples demonstrate the wide-ranging appeal and apparent value of the composite index for communication of results to others (evident in fields as diverse as economics, social analysis, environment, technology and agriculture).

The power of the composite index is its capacity to abbreviate and span. This is also its weakness. The calculation of the final number always requires a considerable number of assumptions about the weighting of components, the relative value of various factors and the exclusion of some items. All of these issues are invisible to the external observer or non-technical person. For this reason, a composite index is always in danger of misuse and misapplication.

One way around the single composite index is to combine various indicators, still in their atomic form, in a scheme or diagram. This enables an overall, visual and readily assessable analysis of a diversity of indicators while at the same time maintaining the independence of each and avoiding the anonymity of factors. The amoeba diagram (see Section 7) is an example of such a device.

3.3 Composite indicators: The example of total factor productivity (TFP)

Although many economists favor the measurement of total factor productivity (TFP) because it takes into account all inputs and outputs, this becomes its Achilles' heel. **CGIAR should focus on partial productivity measures to estimate impact** because these are the simplest measures of agricultural productivity for a single commodity or activity, such as crop yields or output per unit of labor. These are conceptually fairly clear metrics and they are comparable across both time and space, since the units in which input and output are measured are physical quantities (see Box 3.2). **CGIAR should not invest in TFP for impact assessment, although the CRPs may want to measure TFP as part of detailed research studies.**

Box 3.2. TFP and productivity measurement in agriculture

By Doug Gollin

The simplest measures of agricultural productivity are partial productivity measures for a single commodity or activity, such as crop yields or output per unit of labor. They are conceptually clear and are comparable across both time and space, since input and output are measured in physical quantities. The disadvantage of partial productivity measures is that they not useful when productivity is compared across physical locations or moments in time where outputs differ. For example, it is difficult to compare physical units of output per worker in rice and in oranges. For these purposes, economists often aggregate different outputs into common units.

This could still be a physical unit of measure, such as the calorie or protein content of different commodities, but more frequently, economists use production value as a common unit for output.

When prices are used to aggregate output, with no other adjustments, the corresponding productivity measures are gross value per worker or gross value per unit of land. These comparisons are fairly straightforward to compute at a moment in time, but they become problematic once prices are introduced. Because prices differ across time and space, these measures cannot distinguish between changes in quantity and changes in price. To address this problem, economists like to use a common set of prices to value commodity production in different locations and at different dates. These are referred to as 'base' prices, and they are essentially treated as a common set of weights that are used to aggregate quantities across time and space. The resulting measures are 'real' gross values, as opposed to 'nominal' values.

A problem with real measures is that the results may be quite sensitive to the choice of base prices. This is a well-known mathematical result, sometimes referred to as an index number problem. In theoretical terms it means that we cannot be entirely confident in comparisons across time and space. In practical terms, the use of common statistical methods like chain indices is usually considered sufficient for comparisons over time. Comparisons of value across space are somewhat more problematic because they may also involve comparisons between countries with very different patterns of consumption and production, implying very different domestic prices.

The measurement of gross output values also raises another concern. Increases in output typically accompany increases in inputs. These are not free of cost. If output increases solely as a result of input use, then economists do not think of this as productivity growth – simply as intensification.[1] The simplest adjustment for this kind of intensification is to subtract the cost of purchased intermediate inputs from the value of production. Purchased inputs include agricultural chemicals, fuel, seed and feed. They do not include the costs of labor, land or capital, which are viewed as fundamental factors of production. Gross output value minus purchased intermediate inputs equals 'value added,' which economists often use as a measure of output. The concept of value added in agriculture is the measure of agricultural production that is included in GDP. Thus, agricultural value added per worker and agricultural value added per unit of land are probably better partial productivity measures than the corresponding measures of gross output.

Value-added measures are closely related to measures of profitability. They are widely used in economics, but they are probably not very helpful measures of productivity for a research system to monitor. This is because value added can rise or fall due to changes in input as well as output prices. As with gross value measures, it is possible to construct real as well as nominal measures of value added by imposing a base set of prices for inputs and outputs. But in reality the mix of inputs may be very different in different locations or at different moments in time, as a result of changing relative prices. Consider, for example, a government that decides to subsidize the farm price of fertilizer. If the fertilizer is valued at the higher prices that prevailed during the base year, it will appear as a decrease in measured value added.

Value added per worker and value added per hectare are still partial productivity measures, however. Value added per worker will normally increase when there is an increase in land or capital. Similarly, the value added per unit of land will rise when more workers are added. This is why economists particularly like to use TFP measures, which simultaneously account for changes in land, labor and capital. The idea of TFP is straightforward enough. In its simplest form, a TFP measure compares aggregate output to aggregate inputs. Normally, we are interested in changes in TFP over time or across space. For this reason, we use indices of output and inputs. If the index of output grows faster than the index of inputs, we say that TFP has increased.

The indices of inputs and outputs can be computed in many ways. Aggregate output is usually measured in gross value terms. Aggregate inputs can be measured in value terms, but it is more common (and more desirable) to create an index of inputs based on assumptions about their relationship in production. An accepted practice is to use a 'production function' that reflects the underlying relationships between inputs, as estimated through a statistical procedure or as calibrated to data.

For example, a common functional form might relate the log of output to a weighted sum of log inputs. The calculation of TFP will be sensitive to the choice of functional form and the parameterization of the production function, but many economists would view this as the best way to create an index of inputs. It uses physical quantities instead of prices, which removes one source of variation over time and space. However, this does involve an assumption that the same production function applies everywhere, which may be problematic.

Calculating any of these measures requires detailed data on both inputs and outputs. This can be a significant obstacle in any production setting. It is particularly problematic in developing country agriculture, where human labor is a key input; the problem is that labor is difficult to measure with accuracy. In a developing country context, most households divide their labor between farm and non-farm activities. Without a detailed breakdown of the hours worked by family members, it may be difficult to know with confidence whether the labor input has changed from one year to the next. The movement of labor from one season to another, or from one task to another, may also be important and may be difficult to measure confidently.

Another problem is that production itself is subject to high variability. In any given production environment, TFP is likely to fluctuate from year to year because of weather and growing conditions. This will be less important in highly stable production systems, such as irrigated lowland cropping or in animal agriculture involving highly controlled environments. But in marginal areas, such as drylands or semi-arid zones, year-to-year production variability can be very large.

The short-term variation in production environments is mirrored in variation across space. Different locations using the same production technologies will register different levels of TFP. In part, this will reflect differences in the type and importance of production shocks, e.g. one community may receive rainfall at a key moment in the growing season, while another nearby community does not. And in part, these differences in TFP levels will be due to differences – either observable or unobservable – in control variables, such as soil quality, location and access to markets. These differences may lead to persistent gaps in TFP levels, or they may interact in complex ways with year-to-year variation. For example, one community may have soils that are highly productive in good years but poorly drained in bad years; as a result, cross-location differences may be very hard to interpret without a long time series.

Most economists would argue that TFP changes over time are meaningful if they are sufficiently large and persistent. The time scale over which they are measured is crucial. Long-term changes are likely to be more meaningful than short-term fluctuations.

For the purposes of assessing research impacts, the measurement of year-to-year variation in agricultural TFP is foolhardy at best. There is probably no point in assessing research impacts with less than 5 years of data; over this time period, the measurement of TFP is likely to be a wholly unreliable way of assessing research impacts. Over periods of 5–10 years or more, TFP begins to become a defensible – if flawed – indicator of productivity change. Even in this case, changes in TFP will be highly sensitive to start and end points and to weather shocks and the like. In addition, for newly introduced technologies, TFP measures may reflect the difficulty farmers face in learning about new technologies.

Over longer periods of time – 10 years or more – TFP growth becomes a more sensible measure of productivity change. In this time frame, it will be driven less by spurious shocks and measurement error, and it may more plausibly measure a change in technology. Longer periods and larger sample sizes improve the quality of these estimates. Even then, the methodological challenges and data requirements of measuring TFP properly make it a complicated statistical exercise.

Finally, TFP growth creates complicated problems of attribution for agricultural research. Research is not the only source of TFP growth: it could change as a result of policy shifts or alterations in institutional arrangements and incentives; it may be affected by changes in infrastructure and marketing systems. Improvements in the quality of inputs, unless carefully measured, may be confused with productivity gains. For example, a move toward better quality fertilizer – perhaps more appropriately formulated for a specific growing environment – will show up in the data as an increase in TFP, but it should not necessarily be seen as a research impact.

The following are some bottom-line conclusions.

- Measuring TFP properly, whether at the farm level, the landscape level or the national level, is extremely difficult. It requires careful and detailed measurement of all input and output quantities and prices, at an appropriate level. This includes accurate quantification of hard-to-observe inputs, such as family labor time.
- Because agricultural production is highly stochastic, short-term fluctuations in measured TFP, whether positive or negative, cannot be interpreted as changes in the 'true' productivity level.
- In the same vein, cross-section comparisons of TFP levels across locations, at a moment in time, cannot be viewed as representing differences in true productivity.
- Over longer periods of time, changes in TFP can more plausibly reflect productivity change. The length of time needed for confident interpretation is complicated and can be viewed as a kind of power calculation. It will depend on the sample size and the expected variance of productivity within the sample in the cross-section and variance over time. In general, however, calculations based on less than 5 years of data are almost certainly pointless as measures of research impact.
- To measure research impacts through TFP measurement at sentinel sites, as has been proposed in CGIAR, it will be necessary to think carefully about methods and approaches.
  - A near-necessary condition would be to randomly select sentinel sites, so that changes in TFP at the sentinel sites can be compared with changes at the control set of sites.
  - If interventions are undertaken at the community level (i.e. including most or all households within the sentinel sites), the sample needs to consist of large numbers of communities, so that comparisons of treatment and control communities will have sufficient power to give meaningful results over a period of time.
  - Multiple years of data are needed, as well as lengthy time periods. For instance, two waves of data, collected 5 years apart, may not yield convincing evidence of TFP differences associated with research, since production is intrinsically noisy. But with reasonably large samples and reasonably long panels, research impacts on TFP should become apparent.
- Other variables may be easier to collect and to attribute to research. For example, the adoption of new seed varieties or management practices that can be clearly attributed to research are *ipso facto* evidence that farmers value the technologies.
- Measuring TFP badly will be worse than not measuring it at all. Poor measurement or failure to account for the inevitable year-on-year changes may lead to misleading and inaccurate assessments of research impact, whether positive or negative. Decisions made on the basis of misleading assessments of impact will potentially be harmful to CGIAR's mission.

1. In principle, this also applies to output increases that come from increased use of non-purchased inputs, such as soil, organic matter or water.

3.4 Selecting indicators from an understanding of the theory of change

An approach to designing outcome indicators might derive from probabilistic impact pathways. Indicators must be selected with a clear sense of the theory of change or information 'results chain' that links research activities with higher-level outcomes. The notion of the results chain or 'information value chain' is the idea that the results of CRP activities are clearly linked and efficiently provided in real time. While relatively mundane, SMART (specific, measurable, attainable, realistic and time-bound) criteria define the accuracy and availability of data as well as the close ties between indicators and the theory of change used by the CRPs to link their research activities to the IDO and system-level results. One approach to deriving indicators is to do rapid *ex ante* analysis of likely impacts of interventions based on probabilistic pathways, taking into account all known risks and uncertainties (see Box 3.3).

It is important to monitor indicators of the potential adverse effects of CRP research. This could include outputs or activities that may have unforeseen negative ecosystems effects, unforeseen human health effects or negative effects on the bottom of the livelihoods pyramid. Continued assessment of research progress against anticipated theories of change can help monitor such impacts.

Box 3.3 Deriving performance indicators from probabilistic impact pathways

By Eike Luedeling and Keith Shepherd

The research of CGIAR aims to improve lives and landscapes, but some research efforts fall short of this goal. There are many reasons why research may not deliver the intended impacts, ranging from poor research design to inadequate partnerships or sudden shifts in the political environment. Forecasts of research impacts therefore always include a substantial amount of uncertainty. Impact pathways and theories of change help articulate the ways in which research is expected to result in positive changes on the ground. However, these pathways are not normally very good at accounting for risks, which may be noted but are not considered explicitly. They also typically do not consider uncertainties about many other important factors, such as adoption rates or yield benefits of an intervention, which affect the magnitude of the impact that can be expected. Even when impact projections are done in a quantitative way, they often rely on the assumption that all uncertain variables assume best-bet values. This leads to highly precise but possibly very inaccurate projections that mask the risks and uncertainties of the research or development activity in question.

To obtain more robust projections of research impact, projection methods can explicitly include all relevant risks and uncertainties. This requires adding a quantitative dimension to impact pathways or theories of change in such a way that they become functions that convert certain sets of input parameters into quantitative estimates of likely impacts. Examples of such input parameters are estimates of adoption rates or the likelihood that a decision-maker's behavior will change due to information received from researchers. In this process, it is important to be explicit about all the steps that must be taken and events that must occur for impact to materialize. Once such a quantified impact pathway has been established, it can be used to compute probable research impacts given particular sets of input parameters. Most of these parameters are not known with certainty, but their likely ranges and distributions can normally be estimated with some confidence. Once a model is available and input parameter distributions have been estimated, a Monte Carlo analysis can be used to compute the distributions of likely impacts. In a Monte Carlo analysis, a model is run thousands of times with slightly differing combinations of plausible values for all input variables. The result is a probability distribution of likely impacts that allows appraisal of what impacts can be expected and with what level of confidence.

It is also possible to mine data generated in a Monte Carlo analysis with multivariate statistics in order to find out which uncertainties had the greatest bearing on projected impacts. Such an analysis can expose the main reasons why impacts cannot be forecast with greater certainty. These are either pertinent knowledge gaps or risks that can possibly be addressed by modifying the research design or enhancing the effectiveness of actors along the impact pathway. For monitoring purposes, intermediate impact pathway variables – which are outside the researcher's direct sphere of influence but are critical for achieving impact – can also be tracked and tested for their relationship with development impacts. Such performance indicators offer a fairer way to judge research performance than making an evaluation based on impacts alone, because the degree of impact is affected by a host of random factors beyond the researchers' control. The framework described here provides a strategy for identifying intermediate-level indicators that are useful for tracking and anticipating research impact. It also in essence provides a business model for the intended research, presented in a way that donors can judge value for money.

Methods to accomplish what is outlined here are currently in use and under further development at the World Agroforestry Centre, under the umbrella of WLE's Information Systems Strategic Research Portfolio.

3.5 The feasibility of using national or global statistics

Many composite indicators (e.g. the Global Hunger Index) have the advantage that they are available and collected by others. Several of the CRPs have highlighted the importance of selecting indicators together with national partners. This is desirable as the indicators should then be aligned with national-level data and it should be more likely that their collection will be maintained in the long term. However, the feasibility and utility of indicators of this sort should be considered with care when CGIAR is not the only arbiter. This may require a process of iterative scientific debate and CGIAR must always understand the pros and cons of national statistics (as described in Box 3.4 on definitions of forest cover). One reason why rethinking data collecting in CGIAR is so important now is that, for most intents and purposes, data storage is nearly free of cost and large-scale computing has become a cheap commodity.

Harmonizing SLO-level indicators is particularly important to ensure that CGIAR strategically aligns its work with other agencies but is in a position to describe and attribute CGIAR contributions potentially to more local measures.

3.6 Linking metrics and indicators across scales and levels

The difficulties of measurement, and the uncertainties associated with measurements, often increase with the scale at which metrics are measured. Measuring crop yields or soil organic carbon, two of the most basic metrics, is fairly straightforward at the plot or field scale, but when used to derive indicators at higher levels (e.g. at the farm, farm household, livelihood, village, provincial or national levels), it is much more difficult. Measurements can be scaled up by extrapolation, interpolation or aggregation (Volk and Ewert, 2011) – and the most appropriate method depends on the detail of the measurements and the variability in the conditions under which they were taken (Ewert *et al.*, 2011). Scaling up across levels is not simply a matter of describing and accounting for variability, since moving across levels requires that the emergent properties of the systems be accounted for. A simple example is that the integration of nutrient or economic balances across levels as inputs and outputs at one level (e.g. at the level of the crop or herd) become internal flows at farm level, farming system level or above. Thus, great care needs to be taken in specifying protocols for measurements to allow integration at higher levels.

Going beyond simple scaling of metrics and indicators across levels to provide an integrated analysis at different levels is a science in itself. Within CGIAR, there are examples of sophisticated and insightful analyses at different levels – for example, at crop level, landscape level and regional or global levels. To date, CGIAR seems not to have invested in initiatives to link analyses across different levels, from crop to farm to landscape and so on. Integrating metrics across scales will be a major challenge for many of the CRPs – particularly those focusing on large natural resource systems. As trade-offs may exist between goals at different levels, or among different stakeholders, multilevel approaches are needed to ensure coherent and internally consistent analyses. One example of a multilevel analysis system is the SEAMLESS framework designed to support policy decisions in the European Union (Van Ittersum *et al.*, 2008). **The development of a multilevel system that enables the analysis of trade-offs and interactions across system components could allow better integration of results across the different CRPs and was among the gaps and challenges highlighted by the CRPs during the workshop that was convened for this study.**

Box 3.4 Metrics on forests: An example of the difficulties of developing comprehensive metrics for natural resource systems

By Jeffrey Sayer

The extent of forest cover would seem to be a relatively easy thing to measure. One might imagine we could agree on what is and what is not a forest and on how much land it covers. But more than 30 years of attempts to develop broadly acceptable measures of the forest at global, national and local levels reveal that this remains a major challenge. The fundamental problem is that different people value different attributes of forests, and a single metric – forest cover – cannot adequately capture this diversity of values. From this perspective, forests occupy a continuum from agriculture or rangeland with a few scattered trees, through progressively more wooded agroforests, to intense mixed tree–crop systems, intensively managed plantations and disturbed natural forests, through to relatively intact natural forests. The extent, diversity, carbon and biodiversity values of these systems are constantly changing under the influence of natural processes, changing climates and human interventions.

So a line drawn on a map to define a forest will fall in different places depending on whether the person drawing the line is interested in biodiversity values, carbon stocks, timber or non-timber products, watershed values, etc. No single metric for forests will meet the needs of all of these interest groups. For example, Indonesia's deforestation rate over the past 20 years varies between +4 and –0.5 percent per year, depending upon which of seven operational forest definitions are used (van Noordwijk *et al*., 2013).

In the face of these difficulties, FAO has developed international criteria for reporting forest land use and provides regularly updated forest statistics at global and national levels.[1] The criteria are a compromise and represent the best attempt to develop a consensus on the definition of forest. Three sets of problems emerge from this.

- First, the consensus on definitions derives from a political process and changes over time. As a result, the minimum tree canopy cover to qualify as forest has variously been set at 10 percent or been left up to countries to define. When the 10 percent figure is used (as has been the case since 2000), large areas of savannah and steppe are potentially reported as forest. A more flexible approach to definitions, dating from 1948, resulted in reported forest area virtually identical to that reported in 2010, even though the nature of these forests has changed substantially.
- Second, people using the FAO figures may not fully understand the exact attributes of the forest system that is being reported. The FAO global figures combine forest types with low carbon stocks and those with high carbon stocks. They confound biodiversity-rich forests with those that are impoverished. They also represent forest land use – not forest cover. This makes sense for temporarily destocked forest in the same way that a fallowed wheat field is still a wheat field, but it is not the same metric used by others reporting forest cover. Feeding generalized figures into global models of changes in carbon, biodiversity or other forest values may yield misleading results.
- Third, although remote sensing capacity to differentiate different forest types is constantly improving, the ability to conduct the 'ground truthing' needed to exploit technical potential is limited. Feeding generalized remotely sensed map data into local or higher-level decision-making processes is difficult – the capacity to understand and use the data is weak. Likewise, much of the resources can only pragmatically be understood from the ground and the integration of high-quality remote sensing analyses with repeated forest inventory is rare.

The difficulty of deriving generalized metrics is demonstrated by a recent paper on global forest cover using fine-resolution remote sensing (Hansen *et al*., 2013). This paper claims to demonstrate that many prior forest assessments are wrong. In reality, what it shows is that prior assessments may have been partially wrong but also that they used different criteria and methods. The paper has been used to show that the rate of forest loss in Southeast Asia is much faster than had previously been reported. But it is widely accepted that the main driver of forest loss in Southeast Asia is conversion to industrial tree crops. Many of the people who quote this paper have failed to realize that in this case virtually all tree cover is classified as forest, including oil palm and other woody estate crops. Likewise, even-aged forest management in temperate and boreal zones is counted as forest loss rather than part of a forest management cycle. This has led to misleading inferences on the extent and cause of forest loss.

1. FAO. 2010. *Global Forest Resources Assessment 2010* (available at: www.fao.org/forestry/fra/fra2010/en/).

4. CGIAR metrics

4.1 Indicators of IDOs and SLOs, and their relationships

The role of CGIAR is to conduct agricultural research for development. CGIAR research currently aims to contribute to four overarching goals, termed 'system-level outcomes' (SLOs), namely:

1. Reduced rural poverty
2. Improved food security
3. Improved nutrition and health
4. More sustainable management of natural resources (CGIAR, 2010).

The Fund Council found that the 2010 version of CGIAR's SRF failed to make a clear connection between the CRPs and the SLOs. The ISPC's white paper on prioritization made a number of recommendations for strengthening the SRF and filling the gap between the objectives of the high-level SLOs and the research outputs of the CRPs (ISPC, 2012b). The paper recommended the development of a prioritized set of IDOs logically linked to the four SLOs. It also suggested that the CRPs should elaborate CRP-level IDOs to be connected with system-level IDOs, based on robust impact pathways and theories of change. It was argued that this four-fold framework would allow the establishment of coherent linkages between CRP activities and the development goals of CGIAR. It would also enable evidence-based adjustment of CGIAR's research portfolio. The ISPC has contributed a second white paper on SLO impact pathways and interlinkages (ISPC, 2013b), which identifies major ways in which agricultural research could contribute to the four SLOs and the potential linkages between research and impact pathways at the system level.

Thus, the SRF proposes four levels at which CGIAR research could be monitored and assessed (Table 4.1). These four levels share a common basis with the levels of the logical framework approach (Bell, 2000). Starting at the bottom of the hierarchy, measures of performance track outputs to indicate if people are active and timely in delivery of outputs (e.g. breeding new varieties, organizing courses, writing reports). This is largely a management issue and could perhaps benefit from being standardized across the CGIAR Centers, given that many scientists work together on common outputs in the CRPs. In this report, we do not deal with this level of monitoring except in relation to open data management.

Outputs from research are generally tested at local levels (Table 4.1). Monitoring the adoption of interventions provides a basis for the CRP-level IDOs, which are tracked at local levels and can form the basis for extrapolating impacts to the national level. In turn, these contribute to a smaller set of system-level IDOs. These are monitored with impact indicators: measures of the contribution of research to development (*ex post* analysis) or the likelihood of its contribution to development (*ex ante* analysis). Monitoring at these levels allows CRPs to better understand the expected and actual impact of their work. It forms the main tool for research priority-setting.

The system-level IDOs in turn contribute to the achievement of SLOs at international to global levels, which are monitored using a more strategic level of impact indicators, such as absolute and relative changes in poverty among countries.

Moving up through the levels in this hierarchy, it is clear that tracking progress toward the SLOs requires monitoring information, first at research sites and then at country scale – a level for which governments and international bodies such as the World Bank and the UN are responsible (see Table 4.1). The alignment

Table 4.1. The hierarchy of outcome levels in the CGIAR Strategy and Results Framework

Results	Pathway	Levels (scale)[a]	Logframe equivalent	Type of indicator	Role of CGIAR	Monitoring
SLO	Effect of output/ intervention	Global/ regional/national	Goal	Strategic indicators	Assess impact of CGIAR research across all levels from CRP to SLO	Governments, World Bank, FAO, UN
IDO – system level	Adoption of output/ intervention	Regional/ national	Purpose	Sustainability indicators (sometimes known as sustained impact indicators)		CGIAR (across CRPs), national partners
IDO – CRP level	Adoption of output/ intervention	National/ provincial	Outcome	Impact indicators		CRPs, national partners
Output – CRP/ project level	Research outputs	Tested at local level	Activity	Performance indicators		CRP managers, project managers

a The term 'scale' refers to "the spatial, temporal, quantitative, or analytical dimensions used to measure and study any phenomenon," and 'levels' to "the units of analysis that are located at different (hierarchical) positions on a scale" – for a relevant discussion see Cash *et al*. (2006).

of SLO metrics/indicators with a subset of SDG indicators and targets would help illustrate CGIAR's contribution to internationally recognized targets.

4.2 Candidate metrics and indicators for the CGIAR SLOs

CGIAR seeks to establish metrics and indicators that can be used to prioritize investments and to monitor progress. The panel suggests that the contribution of CGIAR toward three of the four SLOs can be quantified with one or a very few indicators. It is important to note that while CGIAR needs data on changes in the SLOs (e.g. changes to poverty levels, natural resource use), the indicators required to assess CGIAR's effectiveness only need to estimate the amount of change that can be attributed to CGIAR research and development.

In the recent CGIAR SRF Management Update 2013–2014, it was emphasized that "No recognized technical methodology can produce a set of metrics that fully expresses the system-level objectives and the causal relationship of changes in those objectives due to uncertain nature of scientific discovery" (CGIAR Consortium, 2013b). We argue that this is a truism. There is no way in which we can represent all of the complexities of the world. Of course, as with anything we do, methods can and should be refined. However, there are already well-established methods for measuring impact, as demonstrated by the large amount of impact assessment literature that is cited in the ISPC white paper. The current debate about approaches to measure impact sometimes seems to overlook this rich experience.

Reducing rural poverty

There is a large and well-established literature on measuring poverty. The most commonly used indicator is the 'headcount,' that is the number of persons living below an arbitrary poverty line. Currently, the commonly accepted poverty line is a purchasing power-adjusted US$1.25 per person per day. The headcount is one of the three standard Foster–Greer–Thorbecke (FGT) metrics (Foster *et al*., 1984). The other two are the poverty gap (income needed to get everyone over the poverty line) and the poverty depth (the gap squared, as a way to give a stronger, non-linear, weight to the very poor). Obviously these indicators are very strongly correlated. CGIAR does not need to show that there are changes in these indicators, or do comprehensive surveys. Instead it needs to show that CGIAR research and development has affected them in a positive way. Changes in measures of inequality such as the Gini coefficient are less relevant to CGIAR's goal to reduce absolute poverty.

While poverty is a multidimensional and complex phenomenon, there is overwhelming evidence that agricultural research can help to alleviate it and there are a large number of studies that provide examples of how the indirect and direct effect of agricultural research on poverty can be estimated (Scobie and Rafael Posada, 1977; Walker, 2000; Hazell and Haddad, 2001; de Janvry and Sadoulet, 2002).

Although this SLO focuses on reducing rural poverty, CGIAR research has been particularly important for consumers, including urban consumers, who benefit from lower prices. Yield-increasing technologies may increase income for the farmer, depending on the effect of higher yields on price. In many smallholder farming systems a majority of the rural poor are also net consumers. Lower prices are always good for the poor as they lead to higher consumption by poor people.

Improving food security

Food security is a complex area that encompasses food availability, access, quality, utilization and stability. The main aspects relevant to CGIAR's work on food security are availability, access and stability, since quality is more linked to the separate SLO on nutrition and health. Availability and access can be summarized as food prices. Agricultural research and development can increase production, leading to lower (local or global) prices, which benefit food security. Another way to express this would be in terms of caloric consumption (Joules per person per day). Variation could be assessed within and between years.

Finally, measures of dietary diversity are increasingly being considered as indicators of food security. Dietary diversity captures both the quantity and quality of the diet. Several specific measures are available that enumerate food groups, frequency of consumption of specific foods and frequency of consumption of food groups. New global initiatives are emerging on monitoring food security, including the FAO Food and Nutrition Security Index (FaNSI) and the Global Food Security Index from the Economist Intelligence Unit (http://foodsecurityindex.eiu.com/), both of which suffer from the common pitfalls of complex indices (see Section 3.2).

Improving nutrition and health

A useful indicator for this SLO is disability-adjusted life years (DALYs). This is a measure of the years of life lost due to premature mortality or disability/morbidity. DALYs are complex measures with many built-in assumptions but they have the advantage of measuring a direct impact. Particular projects or CRPs may also want to express their impact on component changes (e.g. number/percentage of children aged between 1 and 5 years with vitamin A deficiency determined by serum retinol; number/percentage of children under the age of five and women of child-bearing age [15–50 years] who are anemic), although these are only proximal metrics and the relationship with DALYs is assumed.

General improvements in nutritional status, such as the prevalence of stunting among children under the age of five and chronic energy deficiency (CED) among adult non-pregnant women, are simple and standardized measures. They are frequently used in standard age groups, such as children aged 6–59 months. Stunting is likely to be an SDG indicator, but measures closer to the direct outcomes of agricultural research (such as dietary diversity) are more likely candidates for the monitoring and evaluation of CRPs (ISPC, 2013c).

Sustainable management of natural resources

Of the four SLOs, this is clearly the most difficult one in terms of defining indicators. First, we need to define 'sustainable' in a clear way (it is often defined as a synonym for 'good', which is not very helpful).

One approach is to define sustainability as a measure of the rate of depletion of renewable resources. If these resources are not depleted (or if they increase), then a system is sustainable. The higher the rate of depletion, the less sustainable the system is. Whereas some aspects of sustainability can be quantified (e.g. the effect of salinization on crop yield) and expressed in terms of loss of agricultural productivity, the cost of other phenomena (e.g. deforestation and associated species loss) cannot, in our view, be expressed in monetary terms. Although a whole branch of environmental economics is devoted to valuation of ecosystem services, many assumptions have to be made and the values derived are often contested. Thus, monetary valuations are unlikely to be widely accepted. This means that a diversity of indicators will be required for this SLO, which is not a problem, as long as each indicator is clearly defined and measurable. How a particular R&D activity affects this SLO needs a clear impact pathway or 'theory of change', and changes should be quantified.

4.3 Intermediate development outcomes

The role of CGIAR is to assess the impact of its research on development at all levels from the CRP to the SLOs. If a research output, such as a technology, is adopted by farmers, policy-makers or others, this represents a research outcome. For example, when variety X is planted on 100,000 hectares, such an outcome could be used as an indicator of a change that contributes to an IDO.

The Consortium has encouraged the adoption of a set of common IDOs that have been selected to cut across the CRPs (Table 4.2).

The Consortium has suggested that the common IDOs become the building blocks for the proposed CGIAR accountability framework in the SRF (CGIAR Consortium, 2013b). During 2013, two working groups of CGIAR science leaders were established to further improve the definitions of the CRPs and common IDOs, and to initiate the process for developing metrics and indicators for the IDOs and SLOs (Table 4.2 and Annex 7).

Table 4.2. Common set of CRP IDOs

1. Productivity – Improved productivity in pro-poor food systems
2. Food security – Increased and stable access to food commodities by rural and urban poor
3. Nutrition – Improved diet quality of nutritionally vulnerable populations, especially women and children
4. Income – Increased and more equitable income from agricultural and natural resources management and environmental services earned by low-income value chain actors
5. Gender & empowerment – Increased control over resources and participation in decision-making by women and other marginalized groups
6. Capacity to innovate – Increased capacity for innovation within low-income and vulnerable rural communities, allowing them to improve livelihoods
7. Adaptive capacity – Increased capacity in low-income communities to adapt to environmental and economic variability, shocks and longer term changes
8. Policies – More effective policies supporting sustainable, resilient and equitable agricultural and natural resources management developed and adopted by agricultural, conservation and development organizations, national governments and international bodies
9. Environment – Minimized adverse environmental effects of increased production intensification
10. Future options – Greater resilience of agricultural/forest/water-based/mixed-crop, livestock, aquatic systems for enhanced ecosystem services
11. Climate – Increased carbon sequestration and reduction of greenhouse gases through improved agriculture and natural resources management

Source: CGIAR SRF Management Update 2013–2014 (CGIAR Consortium, 2013b).

Generic indicators for the common IDOs are being developed, although the practical difficulty in finding uniform indicators across the CRPs has been noted.

In the panel's view, the system-level IDOs and CRP IDOs do not yet have a real coherence within the SRF. The space between the IDOs and SLOs is quite confused in the current working documents of the CRP and IDO working groups. Conceptually, the system-level IDOs might represent scaled and aggregated outcomes, while the CRP IDOs might reflect smaller-scale and product-specific outcomes. This would give the model greater coherence in relation to the IDO-level components. Similarly, the SLOs will need to include indicators that can be measured in impact evaluation studies at any geographical scale.

The framework can thus be seen as a top-down and bottom-up dynamic. From the top, SLOs are developed in relation to the SDGs. These represent broad geographic scales and make it easy to measure and monitor key indicators. From the bottom, these same indicators appear as impact indicators for studies, and the effect, magnitude and adoption rates can be measured at the CRP level for scale-up through to the system-level IDOs.

The panel believes that it is unclear whether the common IDOs are intended to constitute the system-level IDOs, or whether further system-level IDOs (called SL-IDOs in earlier ISPC papers) will be developed at a 'higher' level. A higher level could imply greater geographic spread (international rather than national level) or be a step closer to the SLOs in terms of the impact pathway or theory of change. The SRF documents suggest that the IDOs should be achievable in 9–12 years, a shorter time frame than the 15–20 years for the SLOs. This implies that achieving the SLOs is an aggregated outcome of the IDOs. At present, no specific target dates have been set for the IDOs or SLOs. The ambition of both is clearly not for CGIAR to totally eliminate poverty and food insecurity. The ambition is to achieve a measurable reduction in the proportion of the population that falls below the poverty line or is food insecure within a defined time frame. As presently conceived, the IDOs and SLOs differ in the timescale for their fulfillment.

An alternative view is that a contribution to an IDO has an immediate impact on an SLO. For example, if a wheat variety reaches a farmer's field and contributes to enhanced productivity it will immediately contribute to both the IDO and SLO. In essence, every change in an IDO must be transformed into a change in one or more SLOs. The impact at the SLO level does not have to wait for aggregated impact at the IDO level (although there may be exceptions). The impacts at the IDO and SLO levels play out over a number of years and can be integrated (with discount rates).

The IDOs can be thought of as proxies for SLOs that are easier to understand and more directly measurable. Although the relationship between the IDOs and SLOs is not necessarily simple and linear, there are certain cases where they are quite direct. If an IDO is zero (e.g. no new varieties in farmers' fields), then the effect on SLOs is also zero. If someone starts consuming orange-fleshed sweet potato their health does not improve immediately, but if it is possible to estimate their vitamin A deficiency and the benefit of the sweet potato, it is possible to estimate the health benefit. In this case, the IDO could be the impact on consumption that contributes to the SLO of improved nutrition. It could be argued that there is a difference in timescale between the achievement of the IDO and the SLO. But is there any basis for the time frames of 9–12 years and 15–20 years? These are hardly different and are relatively far into the future, and choosing time frames of 1–3 years and 15–20 years would seem more sensible. Although variability may make it hard to detect differences over periods of only a few years, activities that have no impact could be weeded out early on. CGIAR has, and needs to maintain, a diverse research portfolio that includes higher-risk upstream research with downstream adaptive research. Obviously, the time frame for impacts in farmers' fields will differ for the development of genetically engineered C4 compared with the delivery time of more incremental breeding goals.

A set of key indicators, such as prevalence of poverty or agricultural productivity, should be measured at all levels from the CRP to IDO to SLO. Other metrics would differ across the levels from CRP to IDO, but when tied together with a robust theory of change can be used to compute an indicator at the SLO level.

4.4 Indicators of 'difficult' criteria

Metrics are required for all major fields of CGIAR activity and the adoption of some IDOs lend themselves to development of quantifiable indicators set against a clear threshold or target. For others, **metrics are poorly developed for fields such as capacity building (in its broadest sense), for certain aspects of NRM, for the functioning of innovation systems, for (social) learning, for empowerment and for 'capacity to innovate**.' For 'capacity to innovate,' the choice of possible indicators is open to debate (see Box 4.1 on measuring innovation). Measurement can be very labor-intensive for some human welfare outcomes, but initial measurement is only a step toward impact assessment. The IDO 'capacity to adapt' seems rather difficult to measure, but we assume it can be done. The important step is to show that this increase in capacity has mattered (or will matter). It may very well be that you can reduce poverty (an SLO) by improving people's capacity to adapt. However, the relevant indicator remains a change in poverty. Capacity to adapt has been an important goal in farmer field schools and integrated pest management. But what mattered in the end was not how much farmers learned, but whether they produced more rice or potatoes, or maintained yields with less pesticide. While some might want to elevate capacity to adapt (or gender equality) to an indicator that is important in its own right, that would not be a reasonable goal at the SLO level for an agricultural research organization.

Thus, while some of these areas may not lend themselves to simple or routine monitoring, more attention is required to evaluate when new metrics are needed and when they are not necessary. For instance, a logical heuristic argument may be a better way of evaluating some IDOs than attempting measurement. The final test will be whether a direct link can be shown between enhancement/achievement of the IDO and impact at the SLO level.

A system for tracking the broader context could be helpful to assist with understanding the contribution of CGIAR research to outcomes. A device such as a regular PESTEL (political, economic, social, technological, environmental and legal) analysis could help disentangle impacts from change that would have happened anyway.

4.5 Alignment with the Sustainable Development Goals

The proposed Sustainable Development Goals for the post-2015 development agenda are currently under discussion. These goals must be linked to targets and indicators that are applicable at the country level but which can also be disaggregated geographically (e.g. to subnational and local levels) and demographically (e.g. by gender). The Sustainable Development Solutions Network (SDSN) has proposed targets and indicators for the six SDGs to which agriculture contributes (SDSN, 2013). Many of the indicators proposed by the SDSN rely on existing monitoring activities, such as those collated by FAO, but there is a clear call for additional investment in data collection and monitoring.

Since the SDGs and the SLOs emerge from the same logic and are different ways of articulating similar ideas, it makes sense to seek alignment between them. This sounds simple, but turns out to be complex as there is no one-to-one relationship between the SDGs and the SLOs (see Annex 9). In addition, the SDGs apply to all countries and cover a wider range of topics than the focus of CGIAR, which means that direct adoption of the SDGs instead of the SLOs does not seem to be a sensible option.

Box 4.1 Measuring innovation and the capacity to innovate

By Krijn Poppe

Innovation is a broad concept. The Organisation for Economic Co-operation and Development (OECD) defines it as the implementation of a new or significantly improved product (good or service) or process, a new marketing method or a new organizational method in business practices, workplace organization or external relations. This implies that innovation activities are all scientific, technological, organizational, financial and commercial steps that actually, or are intended to, lead to the implementation of innovations. Innovation is often linked to businesses, but the public domain can also innovate. This includes the public aspects of agriculture.

The term 'social innovation' has become popular. This concept has at least three meanings. The first points to the need to take the social mechanisms of innovation into account: people have to adapt their working routines to adopt a new method or to make a new product. In the context of rural development, social innovation can refer to the objective of social inclusion. A third meaning refers to social responsibility for innovations: new technologies may have negative aspects for some stakeholder groups, which should be addressed.

Monitoring innovation in agriculture, e.g. for evidence-based policy-making, is not well developed. The food industry and farmers can be questioned directly as to whether they innovate. In Europe, Eurostat's Community Innovation Survey used a questionnaire to survey businesses with more than 10 employees. Countries that have a Farm Accountancy Data Network (FADN, called ARMS in the USA) can include innovation measurement in their surveys and monitoring activities. That would make it possible to relate innovative behavior to the farm's financial capacity to innovate and to link the innovation to outcome indicators, like the income, net value added and sustainability performance of the farm (Van Galen and Poppe, 2013). OECD's so-called *Oslo Manual* (formally *The Measurement of Scientific and Technological Activities, Proposed Guidelines for Collecting and Interpreting Technological Innovation Data*) contains guidelines on collecting data on innovation (OECD and Eurostat, 2005).

In addition to innovation and outcome indicators like value added, yields or sustainability performance (which are all influenced by factors beyond innovation), statisticians have measured aspects of the scientific process that interfere with innovation. Such indicators include the number of patents and research publications on the output side of the knowledge creation activities, as well as R&D spending on the input side.

The lack of data on innovation makes it hard to monitor and manage innovation policies. This has not inhibited economists from judging the efficiency of investments in agricultural R&D by correlating these investments with development in yields or TFP. This type of research shows high rates of return for investments in agricultural R&D – these are mainly realized in the long run, as it takes some time to move new varieties from the lab to the field (Alston *et al*., 2010; Fuglie, 2012).

Capacity to innovate: Monitoring capacity to innovate is even less well developed. At the farm level, FADNs or other types of surveys can investigate bottlenecks to changing farming practices. At the regional level, agricultural knowledge and innovation systems (AKIS) can be mapped and reviewed (EU SCAR, 2012). It has been argued that a well-developed knowledge and innovation system has seven functions (Bergek *et al*., 2010): (i) knowledge development and diffusion; (ii) influence on direction of search and identification of opportunities; (iii) entrepreneurial experimentation and management of risk and uncertainty; (iv) market formation; (v) resource mobilization; (vi) legitimization; and (vii) development of positive externalities.

Innovation systems can be analyzed according to these functions, and mechanisms to develop or improve these functions can be identified. This may call for policy intervention.

The OECD is testing a framework to review the role of government in fostering innovation in the agri-food sector (OECD, 2012). This framework includes an overview of AKIS actors and institutions and a wide range of policies and governance issues. Selected indicators are used to measure efforts, outcomes and impacts. This should allow a country's performance in fostering innovation to be compared to that of other countries. In the OECD test, the following indicators have been suggested.

List of potential indicators of innovation in the OECD's framework to review the role of government policy

Creation or import of new knowledge

Public and private expenditure on agricultural R&D
Number of staff in public and private agricultural R&D
Number of patents registered in the area of biotechnology

Adoption of new knowledge

Public expenditure on agricultural extension and agricultural schools
Number of staff in agricultural extension services
Public and private cost of extension services
Contribution of technological change to TFP
Adoption of specific innovation (e.g. production practices)

Diffusion of knowledge/combination with use of existing knowledge

Contribution of technical efficiency change to TFP
Distribution of farm productivity performance in the sector
Diversification in non-agricultural on-farm activities
Horizontal and vertical integration in the agri-food chain[1]

Enabling market and policy environment to innovate

Linkage between farm support and productivity performance
Entry and exit in the agricultural sector

Induction of innovation

Change in the rate of substitution of inputs
Reflection of R&D demand in public R&D agenda

1. This is often accompanied by transfers of technology and knowledge and can also create the conditions for co-development of new technology and knowledge.

The SRF Management Update 2013–2014 had proposed that a set of around 20 indicators – derived from those proposed for the SDGs – might be selected for impact assessment across the common IDOs (CGIAR Consortium, 2013b). Many of these, such as the indicators related to SDG 1 'End extreme poverty including hunger,' are already regularly reported and analyzed (e.g. through the annual *World Development Report*). The indicators proposed for agricultural productivity are very different from those that have been used in the past.

The crop yield gap is proposed as a more useful indicator than measurement of yield alone. The yield gap is a powerful communications tool. But estimating the yield gap requires that yield be expressed relative to a theoretical potential and therefore we need a clear and agreed definition of potential yield as well as crop modeling to establish what this potential will be (van Ittersum *et al.*, 2013). Should it be the yield gap compared with potential yield or with the water-limited yield in areas where irrigation is not possible? Which crop model should be used? Sustainability is equated with crop nitrogen (N) use efficiency, though high efficiency is often associated with low N fertilizer use, and perhaps is at odds with the indicator on crop yield gaps. Other proposed indicators, such as full-chain N (or phosphorus) use efficiency (percentage), seem very complex and lack established methods for measuring them. Such indicators require a large number of calculations to be made at different steps in the results chain, which will make it difficult to standardize their measurement. The danger is that a 'mass balance' approach will be used with many hidden assumptions. It is hard to understand why such uncharted territory has been chosen as the basis for an indicator at this scale. What is a 'full chain'? How do we aggregate – or sum up – across the chain? Most pollution is due to the release of organic waste from animals or, more commonly in developing countries, from urban areas, not from agriculture. Thus, such indicators seem far too complex to use in such a general way.

It is also unclear where the responsibility for measuring these indicators will lie, and whether this will be done annually or at less frequent intervals. Monitoring of some – such as crop yields – should best be done annually, because there can be large inter-annual variability, and because both the trend and the inter-annual variability are of interest. Others are probably best measured every few years (e.g. health and poverty indicators).

As discussed in relation to TFP (Box 3.2), there are strong arguments for selecting indicators that are simple and robust, that involve the fewest hidden assumptions and that are easy to understand and communicate to a wide audience. The indicators selected for the SDGs so far do not seem to meet these criteria.

4.6 Foresight

An information system is by definition oriented toward the retrospective selection, accumulation, storage and representation of data. These are often cited as the key features of data-based systems and may constrain the design considerations of the authors of such systems. The latest CGIAR SRF Management Update notes the value of including "forward-looking, dynamic and foresight dimensions in the SRF" (CGIAR Consortium, 2013b). This sentiment needs to be echoed in the design of the CGIAR metrics system. Considerable effort is being devoted to compiling historical data. Once this has been achieved, an almost instinctive reaction of users is to ask 'what if' questions. Historical trends beg future projection. **Cause and effect relationships derived from studies of past trends naturally result in requests for exploration of future effects.** To retrofit a foresight component to a system is often more difficult and costly than to include it in the original design. Foresight, often referred to as 'scenario planning,' includes elements of:

- asking 'what if' questions;
- having the ability to model future trends;
- having a means to identify the causal links between indicators.

Making provision for such foresight modeling approaches in the development of the metrics system will provide highly useful capacities in the information system.

4.7 Summary

In the design of a metrics and indicator system, CGIAR needs to decide whether this should be an 'open' system from which external users can generate reports or a more closed system for use by CGIAR only. If an open system is chosen, it will be necessary to design an intuitive and easy-to-use interface. Presentation devices such as the amoeba diagram (see Section 7) have the advantage of being transparent and easily understood.

Is a single, unifying system of metrics across CGIAR desirable and achievable? The current bottom-up approach is enriched by the experience of realities on the ground and the engagement of local and national partners. A parallel effort is needed to ensure consistency in approaches across the CRPs, which will make synthesis and cross-comparison possible.

Integrative and cost-effective metrics are required for monitoring progress, which would assist in comparative management and decision-making, and to communicate advances in achieving the CRP targets. Comparative analysis across the CRPs and their many project locations is a powerful tool for understanding the context across a hierarchy of levels from local to global. This can only be achieved through strong efforts to archive all past and current CGIAR research.

5. Data management

5.1 Emerging trends in agricultural research and data

Attitudes about managing public research data have shifted considerably over the past two decades. It is now generally assumed that the raw data emerging from research will be made available as part of scientific publications. Publishing in peer-reviewed journals allows for better scrutiny and reproducibility and, more importantly, it enables further research, whether through an alternative analysis of the same data or by combining it with other data. Nevertheless, despite the expectation of data availability and the explicit requirements to this effect from leading scientific journals, most publications are currently not accompanied by the relevant raw data, in part because (until recently) the informatics infrastructure was lacking. This is no longer the case.

5.2 Big data

A major recent development is the availability of very large data sets and new analytical tools (e.g. machine learning algorithms) to analyze them. New sources of data include satellite or ground-based sensors, DNA sequencing machines, Internet searches and crowdsourcing. While there will always be a need for highly controlled experimental work in which the high quality and specificity of data is crucial, there are many new opportunities where the amount of data is more important than the quality of a particular data point or knowledge of the purpose for which the data were originally collected.

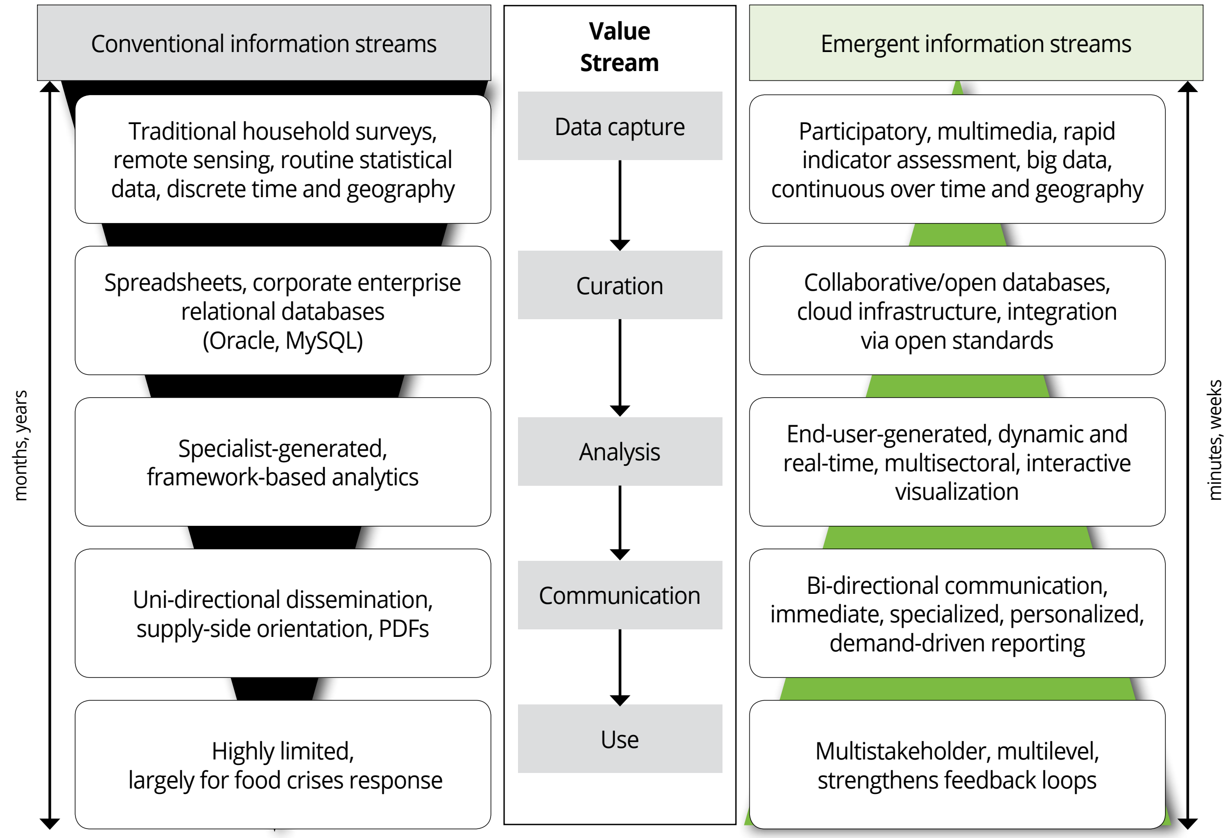
This may be particularly true for agricultural development. The site and time specificity of agriculture make it difficult to understand much of its complexity from small data sets. Through the accumulation of large spatio-temporal databases on aspects of economics and health, agricultural production practices and ecosystem services, we may be able to gain a much better understanding of the dynamics of agricultural change, its sustainability, its influence on the well-being of people, the role that CGIAR has played in the past and the role it can play in the future.

Mock *et al*. (2013) developed a model of the information value stream that ties data, metrics and learning systems together (see Figure 5.1), which is instructive for CGIAR. The information value stream is the set of activities linking an information project to its ultimate use, including as a support for decision-making. CGIAR should take advantage of the potential that information and communications technology (ICT) tools and systems thinking now offer to create a learning system that is driven by evidence and can be used by a wide and diverse stakeholder body.

With the growth in information technology, the intersection of systems approaches opens up possibilities for more dynamic metrics and indicator systems, which could foster organizational, cross-organizational and multistakeholder learning, as illustrated by the emergent information value streams. Here, new data sources including 'big data' – large data streams – can be tapped into. These data sources are increasingly available due to mobile computing, digital technologies and increasing bandwidth throughout the world. There are four principal types of big data:

1. data exhaust, or data that is collected electronically as a function of some other transaction, such as cell phone use;
2. physical sensors, such as environmental monitoring;
3. citizen reporting or crowdsourcing;
4. Internet data (Letouzé, 2012).

Figure 5.1. Information value stream in relation to SRF and the CGIAR learning system.



Big data is particularly useful due to its wide coverage and continuous nature, allowing monitoring over time and often enabling near real-time access. Big data are often unstructured and may include multimedia data, such as images, text and sound. Crowdsourcing is a new way for a stakeholder anywhere in the system to provide information through the value stream. These new data sources are giving rise to new sources of metrics and indicators (e.g. anthropometric indices based on image data, monitoring household liquidity through a proxy of cell phone air card purchasing increments); increasingly simple and comprehensive biomarker assessments; as well as to new metrics (e.g. variance metrics associated with a number of ecological variables). The definition of these new metrics and their validation is in its infancy, however. Several research projects are developing applications (UN Global Pulse; www.unglobalpulse.org) of potential interest to CGIAR, including:

- food price crisis monitoring
- online content monitoring for generating insights on women and employment
- global legal timber trade.

CGIAR will undoubtedly develop many new data sources and analytics in its areas of comparative advantage (agricultural production, soil health, nutritional status). CGIAR's broader Open Access/Open Data initiative will facilitate the movement of traditional information sources to the more modern information value stream, which also will facilitate development of these more novel forms of information.

The emerging field of visualization analytics democratizes the process of data analysis, allowing various stakeholder groups with Internet access to easily access data on demand, often free of charge. Combined with open access data standards, the availability of these tools permits CRPs to conduct a wider range of exploratory analysis. Moreover, when open access is combined with freely available analytical tools, expanded users/stakeholder groups participate in the metrics and indicators learning system. For example, Digital Green (http://www.digitalgreen.org) is bringing dashboards of indicators to farmers through an off-line/online platform that they developed as part of their peer-to-peer farmer capacity development efforts.

5.3 Meta analysis, data annotation and potentials for interaction between data sets

Given the breadth of disciplines represented in CGIAR, the data generated can range from genealogies of varieties, results of agronomic trials, soil and plant analyses, occurrence and spread of livestock diseases to survey data and notes from interviews. Local testing of theories or hypotheses may involve small data sets, whereas building and testing more general theories typically involves larger comparative data sets that allow an exploration of the influence of context, for example the importance of the agroecological and socioeconomic environments on outcomes. Testing the degree to which theories, hypotheses or patterns of response are generally applicable can be done through meta-analysis. In turn, meta-analysis is possible only when many observations or tests of a particular theory are available, which explains why meta-analysis has only recently become widespread. Meta-analysis is still often hampered by lack of access to original data and/or full description of statistical tests performed.

Research in the CRPs is embedded in specific theories of change or impact pathways to which all research activities are expected to contribute. Research results tend to be local in nature (with the exception of broader regional or global studies), which means that results need to be situated within a hierarchy or network of levels to allow the results to be scaled. Thus, crop or animal yields at the field level need to be situated in farms and households, in farming or land-use systems, in regions, countries and so on, to allow the estimation of the impact at higher levels. A recent ISPC-commissioned study highlighted the need to consider farm size and access to markets as criteria for prioritizing research investment, noting the need for special attention to the so-called 'hinterlands' (Masters, 2013). The Global Environmental Change and Food Systems (GECAFS; www.gecafs.org) project highlighted the links between food security and global environmental change whereby demands for food and other products are often met through agricultural activities on different continents. This emphasizes the need for data to be annotated and embedded in their (thoroughly described) context to allow more general analysis and conclusions.

Even more traditional data collection tools, such as population probability sampling, are used so that estimates of nutrition and health, for example, can be disaggregated to lower administrative and geospatial units of scale, emphasizing greater population coverage (and larger sample sizes) on a smaller number of meaningful indicators. **CGIAR Research Programs and Centers can contribute by conducting analyses that yield more parsimonious and accurate metrics related to yield, food security and dietary intake, for example.** As global learning about sustainable food and nutrition security deepens, the trend of combining simple classical indicators with real-time measures of variation, stability and change will increase. **Combining a small set of relatively standard SLO indicators with more real-time data on key drivers of change and/or intermediate results outcomes might be a promising strategy.**

The challenges of measuring forest cover (see Box 3.4) apply equally to the measurement of many of the attributes of the agricultural systems in which CGIAR works. It is not easy to use satellites to determine what rice variety is grown where or how many people have access to a community forest. Big data works only if you have a lot of relevant (field) data. What data does CGIAR have that can be made available? And what new data need to be collected such that we start building up the capacity to better understand the dynamics of agriculture and NRM in developing countries, how research and policy have influenced these processes, and how they may influence them in the future?

5.4 What has been done: Success stories

The CGIAR Centers have led high-profile international data-sharing activities for a long time. Examples include FishBase (fishbase.org) and ReefBase (reefbase.org), which are operated by WorldFish. ICRAF manages the Agroforestree Database, which includes information on the management, use and ecology of a wide range of tree species that can be used in agroforestry (worldagroforestrycentre.org/resources/databases/agroforestree). CGIAR scientists have also made available some of CGIAR's most widely used spatial databases, such as the improved spatial database on elevation (SRTM elevation) (srtm.csi.cgiar.org) and the WorldClim global climate database (worldclim.org). ICRISAT's longitudinal Village Level Studies, a truly one-of-a-kind data set for understanding agricultural development in South Asia, is now also available online (vdsa.icrisat.ac.in).

HarvestChoice (harvestchoice.org) has developed a landscape-scale evaluation framework to organize key agricultural data layers into a standardized matrix of 10 km x 10 km grid cells across Sub-Saharan Africa. This platform allows visualization and examination of the mix of farming, cultural and socioeconomic conditions, by compiling data sets on various biophysical and socioeconomic parameters, including characteristics of soil and climate, market access, farm production systems (area, yield and production of major food crops), potential distribution and persistence of major crop and livestock pests and diseases, characteristics of farm households and the incidence and severity of poverty.

All of these CGIAR data products come from relatively centralized projects that are typically managed by a single Center with dedicated staff providing continuity. As such, they represent relatively simple efforts in terms of management, but still require a large investment. Also, these are not examples of research databases in the sense that they include raw data generated by a particular research activity. Rather, they were data projects, whose purpose was to develop a database for others to use, as an input to research and development activities. There are probably several other data sets that could be developed and/or made available and could become equally important.

A different model is Genesys (genesys-pgr.org). This database provides access to genebank data from CGIAR Centers, but also from the United States Department of Agriculture (USDA), European genebanks and others. Genesys evolved from the System-wide Information Network for Genetic Resources (SINGER), which held data from the CGIAR genebanks. SINGER was, in the early 1990s, one of the first large federated databases on the Internet. **It was a rare example of a group of CGIAR researchers coming together (through the former System-wide Genetic Resources Program) to standardize, coordinate and improve data management practices. The success of this group should be taken as an indication of the possibilities for consolidating data, even if they started out with huge data sets that were being actively managed.** The amount of time and effort it took to build SINGER should be a warning. It is not that difficult to archive raw data, but creating a federated information system can require major investments and a long-term commitment. The CGIAR Consortium for Spatial Information (CSI) has attempted to combine the spatial geographic information system (GIS) data resources of different CGIAR Centers; however, they have not yet been as successful.

5.5 Opportunities for providing greater access to research data

CGIAR has been much less active in making available primary research data from, for example, experimental trials (agtrials.org is an attempt to address this), farm surveys or the phenotyping associated with molecular breeding (Zamir, 2013). This is a much harder task because it involves very heterogeneous data generated by hundreds of researchers. A number of Centers, including IFPRI, IRRI and ICRISAT, have started to publish such data sets through the Harvard Dataverse network (e.g. exploreit.icrisat.org;

thedata.harvard.edu/dvn/dv/IFPRI), a repository for research data for long-term preservation that allows researchers to share, control and get recognition for their data. An example clearly illustrating the value of such work is IFPRI's data set 'Chronic Poverty and Long Term Impact Study in Bangladesh,' which has been cited more than 20 times since its release in 2010.

CGIAR is also a partner in the Coherence in Information for Agricultural Research for Development (CIARD) movement (www.ciard.net/), which appears to be primarily focused on advocating open access to agricultural information.

Other initiatives in which CGIAR does not appear to be involved include the Research Data Alliance. This is a broad initiative by the governments of Australia, the EU and the USA to enable open sharing of data. There are also a number of relevant special interest groups on data citation, data description, biodiversity and big data analysis, in addition to the Agricultural Data Interoperability Interest Group (https://rd-alliance.org/internal-groups/agricultural-data-interoperability-ig.html), which currently focuses on genetic information on wheat.

5.6 CGIAR developments in open access and data management

Not only is data management infrastructure available, but the expectation of open access has also gained considerable ground in the research and practice community. Sharing research data is now officially required by CGIAR and the *CGIAR Guidelines for Open Access* were endorsed while this report was in preparation. The CGIAR Open Access and Data Management Policy (adopted in October 2013) states that all research data should be made available within 12 months of collection (see Box 5.1). The Consortium has also made proposals to the Fund Council for an implementation plan for open access and data management. Although there are still details to be worked out, and the plans focus on future rather than historical data, these are encouraging signs. The rest of this section considers some of the elements for a successful data management system – some of which have been mentioned in the Consortium's implementation proposal, but are still to be developed by CGIAR.

CGIAR will make available a large data resource through its new open access and data management initiatives, enabling substantial improvement in its information value stream and a larger reach to stakeholder groups in and outside the CGIAR Centers. **There is great scope to improve the information value stream by mining existing CGIAR data and experimenting with novel data streams.** Moreover, the rapid development of data science will allow CGIAR to identify and use novel methods to generate more useful and dynamic metrics and indicators in future.

The adoption and implementation of the policy is a landmark decision with potentially tremendous benefits for global agricultural development, not only because of the value of the data produced by CGIAR, but also because this type of leadership is likely to induce others, such as national agricultural research institutes and universities, to follow suit. It will also directly benefit the Centers themselves, as CGIAR's primary research data are often not even available to CGIAR staff.

The panel welcomes the measures proposed to implement the Open Access and Data Management Policy. A degree of centralization of data management, and especially archiving, will be needed but the panel favors a distributed network of data hubs linked with minimal superstructure (e.g. a common portal) but with quality control and curation. Strong leadership and incentives will be needed to ensure widespread respect for and adherence to the policy. This is in part because publishing data demands time and other resources for organizing and curating data. These resources need to be allocated carefully. Fortunately, if computer systems are well designed and supported, data publishing, once it has become routine, could save time and resources and thus easily pay for itself. Several CGIAR Centers already

Box 5.1 Open access initiatives for agricultural data and CGIAR

CGIAR has joined the Global Open Data for Agriculture and Nutrition (GODAN) initiative (www.godan.info). This was officially announced at the Open Government Partnership (OGP) conference on 31 October 2013. The GODAN initiative "seeks to support global efforts to make agricultural and nutritionally relevant data available, accessible and usable for unrestricted use worldwide. The initiative focuses on building high-level policy and public and private institutional support for open data. The initiative encourages collaboration and cooperation among existing agriculture and open data activities, without duplication, and brings together all stakeholders to solve long-standing global problems." GODAN is an initiative of the UK Department for International Development (DFID), the United States Agency for International Development (USAID), the Bill & Melinda Gates Foundation and the World Wide Web Foundation, and it has an impressive membership. Precisely what GODAN will do is up to the participating organizations and CGIAR could play an important role in shaping the initiative.

have data management units that are well equipped to provide the necessary support. It will be critical to ensure that this historic opportunity to change how agricultural research is done is embraced throughout the CGIAR system.

Having standards for data description and ontologies at the time of collecting would obviate the need for describing data at a later stage. But the greatest hurdle could be the lack of commitment from individual researchers. In the past, there has clearly been a reluctance to share data, and it is easy to offer reasons for not doing so. For example, in a recent white paper about data sharing by 22 CGIAR researchers – including at least one from each Center – it was stated that "from a scientific point of view, it is not desirable or useful to make the data from long term trials available" before they have been analyzed (Gassner *et al*., 2010). It was also suggested that it would be reasonable to make economics data available "5–10 years after the data were collected" and bioinformatics data at an unspecified "appropriate time." It was further stated that data collected in the context of research by graduate students, "can only be publically released after papers using the data have been published" and that "it may not be worth the effort to publish 'base-line' data because their quality tends to be too low to be useful." The authors did not explain why such low-quality data are being collected in the first place. The motivation for individual researchers to keep data to themselves is often based on an unjustified fear of being 'scooped.' But not sharing data puts the interests of individual researchers above those of their Centers, CGIAR and its mission. The general rule should be that data will be published within a year of being collected and this policy should be enforced, allowing for exceptions with automatic expiration dates, e.g. for projects that can only yield results after several years of data collection.

CGIAR must urgently consider how to stimulate a research environment where data publishing is expected, stimulated and, where necessary, enforced. It also needs an informatics infrastructure that makes it easy to make the data easily accessible and usable by others. As Gassner *et al.* point out, this requires that data management be included in project budgets and that it is important in performance evaluation. Gassner *et al.* also discuss issues of attribution and ownership as potential barriers, but this problem can easily be overstated. The opportunity for career development through publication is an important aspect, but scientists only succeed if they publish their results in a timely manner. Published data sets only increase the impact of researchers on the scientific community and thus their standing in their respective fields. Recently, there has been marked progress toward publishing more data sets in the field of agronomy (White and van Evert, 2008). A new journal, *Open Data Journal for Agricultural Research* (www.odjar.org), has been launched specifically for that purpose.

The evaluation of research projects should not only consider the availability of data, but also its quality, completeness and ease of use. It is easy to overlook the amount of work it takes to produce high-quality and easy-to-use databases. For example, many data sets that are currently made available by CGIAR are difficult to access because of the need to register, use passwords or even to send email requests to individual researchers (who may no longer work at the Center and/or may not reply).

CGIAR needs to consider the extent to which data sets should be accessed singly or whether action should be taken to allow integration. The AgTrials database is an attempt to make available agricultural trial data. Unfortunately, AgTrials does not allow all data to be downloaded at once – and for some data sets access needs to be requested via email to the original data provider. This defeats the purpose of data aggregation. Also, an inspection of AgTrials illustrated why data curation is a tedious and difficult task. The data for many experiments is incomplete. For example, there are data sets with treatments '1,' '2,' '3' and '4' without explanation of what these treatments entailed. Most experiments are not accompanied by detailed (daily) weather data. AgTrials has not attempted to standardize data due to the large amount of time this would take, but the database serves the purpose of highlighting what data are available and providing a means to access the raw data. Standardization would allow for much easier use, and would help with the design of future studies involving data collection fieldwork. Such initiatives deserve significant investment to create a resource for future research.

5.7 Creating systematic data sets

Given the current interest in the development of comprehensive approaches to data collecting and archiving, a fundamental question remains: who will invest in data in the long term? While most projects budget for the costs of data collection and analysis, few include budget lines for archiving and storing data – yet this is a time-consuming and expensive process. Costs can be streamlined by using standard and agreed formats for data entry and nomenclature. Initiatives to establish simple standards need to be prioritized.

Speaking a common language – The need for an ontology

A major hurdle is the lack of a common ontology of concepts and vocabulary. This goes beyond the need for a simple glossary, as an ontology needs to encompass the way the concept is used. Terms that are commonly used in household surveys, such as crop yield or crop area, cannot be compared if they are not specified accurately (see Box 3.1). Often, indicator compendiums are developed to contain **complete descriptions of indicators. An indicator compendium for the SLOs and system-level IDOs would be advisable. The terms used for geographic locations (sentinel sites, action areas, hubs) needs to be harmonized.**

The European research framework – the SEAMLESS project[3] – invested considerable effort to integrate a wide range of data that were dispersed among different databases and institutions. This required researchers "to ensure consistency in data interpretations, units, spatial and temporal scales, to respect legal regulations of privacy, ownership and copyright, and to enable easy dissemination of data" (Janssen *et al*., 2009, 2011). The project was relatively small in relation to the size of the CRPs, although it involved some 150 scientists from more than 30 research institutes and universities in Europe, together with collaborators from Australia, Mali and the USA. Nevertheless, it was found necessary to develop a shared ontology of agricultural concepts as part of the research process. The ontology goes beyond defining specific concepts

3. The SEAMLESS project (www.seamlessassociation.org) developed an integrated framework for integrated assessments based on linking individual components (models, data, indicators), which enables analyses of the environmental, economic and social contributions of a multifunctional agriculture and the effects of a broad range of issues (e.g. climate change, new policies, innovation).

to define the relationships between the concepts. Confusion over concepts and relationships between concepts will always emerge in multidisciplinary projects. This is not helped by the manner in which implicit segregated ontologies often do not even recognize that they use the same concepts under varying use frameworks.

The FAO Division on Agriculture Information Management Standards (aims.fao.org) maintains the AGROVOC vocabulary (aims.fao.org/standards/agrovoc/about). This system is strong on the ontology of different terms but does not go into the detail of measurement standards and units needed to support consistent measurement.

Survey data could be much easier to integrate if standard questions and coding were used for at least a common core set of questions (and additional more specific modules could also be developed). A successful example of such an approach is the Demographic and Health Surveys (www.measureDHS.com) that have been carried out in 89 developing countries over the past 20 years (the mean number of times each country has been surveyed is 2.9). Again, the intended use of the data should determine the way data are collected in the first place.

5.8 Implications for data collection, collation and storage

Given the enormous range of CGIAR research, any data management system needs to be sufficiently flexible to allow for all needs. Virtually all data processing and calculations involve assumptions about data attributes, and advances in scientific understanding may arise from revisiting these assumptions. To allow future integrative research to make full use of past investments, the data need to be stored in basic, building-block form. A clear example of the pitfalls of research arising from multiple assumptions is given on the estimation of TFP (see Box 3.2). This and other examples lead us to conclude that it is best to carefully measure and store basic data – crop yields, area cropped, etc. – and to ensure that a full description of the methods used for data collection is made available along with the data. This will create a really valuable resource for future research. It will allow any number of analyses using new techniques and assumptions far beyond the data normally presented in scientific publications and reports. When combined with monitoring, using repeated measurements, this will allow development of time series and reveal trends of change.

5.9 The ethics of data use

There are ethical as well as technical concerns that arise with the use of data. Ethically, protection issues can arise when data relating to identity and geographic location are widely available to a broad user base. This issue has been raised, for example, in the context of vulnerable children who make their presence known during emergencies (Morrow *et al*., 2011). Technically, the representativeness of information can be questionable, depending on the population or area coverage of the data stream. For example, SMS data streams may not reflect vulnerable populations in areas of the world with low cell phone coverage, although cell phone coverage is expanding globally. Crowdsourced data reflect frequency counts as opposed to denominator-based measures. Data exhaust originates from electronic transactions, which may not be widespread among vulnerable populations in low-income countries. On the other hand, the ability to have wide geographic coverage at low cost may outweigh these considerations.

An additional technical challenge is data availability. While a large number of publically available databases exist, the private sector owns many of the most useful databases, such as mobile phone, health and banking data. Appropriate standards for data curation and use will have to be developed for CGIAR, including for data provided by our partners.[4]

4. See previous ISPC discussion of ethical considerations in CGIAR (CGIAR Science Council, 2008).

6. Monitoring – Roles and responsibilities

It is clear that CGIAR cannot operate efficiently without a clear understanding of what can be done, what works and what does not. To learn from its innovations and assess impact, CGIAR needs to have access to relevant data at the global level, or at least at the national level for CGIAR's target countries. While it would be beneficial to have new and specifically collected data for monitoring agricultural change, how much should CGIAR invest in this? Another key question concerns the extent to which monitoring agricultural change is the responsibility of CGIAR.

6.1 Data needs and responsibilities

The effect of CGIAR on SLOs can be estimated with conventional methods of impact assessment. This includes estimates of the contribution of past activities to the SLOs (*ex post* analysis) and prognostic studies that estimate the likely contribution of proposed activities to the SLOs (*ex ante* analysis). This type of impact assessment requires measuring the IDOs as well as a number of other variables (e.g. Who adopted the variety? What was the effect on production costs, yield or health? What is the effect on income of producers and price for consumers?). Impact assessment is an important tool for research priority-setting. Data are needed on where certain crops are grown, the sizes of farms in different locations, income, health, etc. The responsibility for collecting this type of data lies mainly with national governments, with support from international bodies, such as the World Bank and FAO. While it is the role of governments to collect comprehensive statistics (e.g. through population, household and agricultural censuses), there is a big task to be done in data aggregation across countries. FAO and the World Bank already do some of this, but such work needs to be expanded. For example, there is no comprehensive source of crop production data (area, yield, over time) at the subnational level. Governments do not always collect primary data (or make them available), and the quality of such data is often poor (Jerven, 2013). While collecting such data is not the true responsibility of CGIAR, they are needed for analysis and priority-setting and have importance far beyond CGIAR. So, should CGIAR step in and collect primary data in the countries where it is conducting research? Or can governments and FAO be persuaded to do more? CGIAR could work with FAO to compile subnational livestock and crop area and production statistics and could enhance these statistics both spatially and temporally through remote sensing.

For attribution or analysis of its contribution to CGIAR research outcomes to be possible, much more specific work will be needed in most cases. Measuring progress toward specific goals and impacts on productivity, sustainability, health or income will inevitably require extensive fieldwork designed to evaluate particular technologies or interventions. **A key strategic issue that needs further analysis is precisely what CGIAR needs to monitor and where, and to what extent it can rely on monitoring by others.** In an ideal world, CGIAR would be able to tap into robust national and subnational statistics on agricultural production and trade that it could then combine with specific indicators to track progress. Unfortunately this is not the case in many of the CGIAR target countries.

An alternative view is that CGIAR should continue to track the big picture, to collect evidence that CGIAR research products (be it agricultural technology or policy advice) are being used and then estimate their effect through case studies of impact assessment based on broader monitoring. Impact assessment should recognize the role of serendipity. Portfolio approaches are needed to avoid the past tendency to focus impact assessment on those innovations that were perceived to have worked; we need to learn from failures as well as from successes. It is hard to predict which research investments lead to most impact. CGIAR culture must allow CRPs to focus on the research and hire the best researchers, spread the products, and then conduct honest *ex post* assessments of what worked and what did not. An example of what worked is to be seen in CIAT's actions on the biological control of the cassava mealybug in Africa (valued at US$9 billion, far exceeding CGIAR's total investment in Africa). CIAT was able to achieve major impact because it provided stable employment to a competent entomologist who was able to find the pest's natural enemy in South America. This work did not emerge from *ex ante* impact studies. The very nature of research means that success is difficult to predict and that it can benefit from relative freedom to explore. Research institutes have to maintain focus and for that reason monitoring is necessary. But is also very easy to suppress progress by imposing unnecessarily complicated monitoring systems. In the end, it is not important whether an institute delivers what it promised. It is important that it delivers something of high value.

It is only with hindsight that one knows which metric would have been the most useful for measuring a particular change. On the one hand, it is important to foster the collection of generally useful data about agriculture and natural resources use, human health and poverty. Such data will be useful no matter where they are collected, and can mostly be left to others (FAO, World Bank, universities). A role of CGIAR could be to encourage better data collecting and monitoring by others. On the other hand, there may be a need to collect highly specific data in particular agroecosystems as they relate to particular research goals (e.g. the current extent of no-till rice–wheat systems in India). New technology, such as remote sensing, could help in some cases. For example, IRRI is using radar satellite images to improve rice mapping (and production statistics).

6.2 Baseline studies

CGIAR runs the risk of devoting substantial resources to measuring and monitoring data that may never be used. The need to demonstrate impact has led researchers to include extensive baseline surveys in most large projects, such as AfricaRISING and many of the projects funded by the Bill & Melinda Gates Foundation. The extent to which these baseline surveys will eventually be analyzed and used in impact assessment is unclear. This is an example of how a greater degree of standardization would add enormous value and cost efficiency. The information collected from the many thousands of surveys conducted in Africa over the past 5 years would provide a rich picture of the distribution of farm sizes and many other attributes of agricultural systems. Gassner *et al*. (2010) also question the usefulness of baseline surveys based on concerns as to their quality. We concur with this view, but if quality cannot be ensured then why are the surveys conducted in the first place?

Where CGIAR does invest in surveys for baselines or other purposes, the efforts need to be more systematic and to harness thematic and geographic complementarities, including biophysical and socioeconomic characterization and analysis (e.g. value chains, policies, adoption and impacts, livelihoods and vulnerability). **CRPs should also fully exploit existing opportunities and economies of scale in collecting and sharing data. This can also help to prevent interview fatigue in the target communities. Standardized methods and measures need to be used for survey and data collection. More thought needs to be given to the measurement of indicators at sentinel sites, hubs, benchmark sites, etc. Harmonized approaches are needed and the ontology of terms for these geographic locations needs further elaboration. The value of the data sets derived from baseline studies will lie partly in their value for assessing impact but more in their value for understanding the processes of change and hence the opportunities for research.**

6.3 Other initiatives on monitoring agriculture, land use, health and poverty

The need for better metrics is not unique to CGIAR and there is the possibility to capitalize on the programs and experiences of others. There are a number of international institutes, partners and donors that collect metrics and indicator data at different spatio-temporal scales and for different purposes. Donor agencies, such as the Bill & Melinda Gates Foundation, USAID and DFID, have launched several initiatives and projects aiming at the development of new metrics systems for pre-assessing and monitoring adoption outcomes for agricultural technologies to determine how investment targets can be achieved, while avoiding potentially unforeseen adverse environmental impacts. A recent study by Shepherd *et al*. (2013) reviewed the current situation with metrics, drawing on the literature and activities of select international initiatives to set up data monitoring systems for agriculture, ecosystems and/or poverty. The study drew useful lessons from measurement and monitoring systems in the public health sector and from the field of decision analysis. The study emphasized the need for a "decision analytic conceptual framework" to enhance the relevance and cost-effectiveness of metrics systems by linking measurements to decisions.

Building on the Shepherd *et al*. study, we collated information on related projects and initiatives on metrics and indicators (see Annexes 5 and 6). This inventory is no doubt incomplete but highlights the large number of ongoing initiatives that could contribute useful data. **The main lessons learned from the review were that the relevance of research for sustainable development is ensured by the contextualization of research and a dialogue between policy and science.**

The key question is whether this information can contribute to CGIAR's four SLOs: reducing rural poverty, improving food security, improving nutrition and health, and sustainable management of natural resources. Will these initiatives contribute current spatio-temporal data on rural poverty, food security, nutrition and health, and sustainability of NRM? Where there is clear thematic and geographic overlap, CGIAR should engage in opportunities for collaboration and joint learning. Criteria for selecting partnerships are needed to ensure that CGIAR remains focused on the SLOs and does not get drawn into trying to measure everything everywhere.

6.4 Use of metrics and indicators for accountability

CGIAR gained considerable experience from the Performance Measurement System that was implemented for 6 years, ending in 2010. Immonen and Cooksy (2013) conducted a detailed evaluation of the Performance Management System that was "intended to become part of a streamlined monitoring and evaluation system and to enhance transparency, accountability, learning, and decision making, including decisions about future funding" (Immonen and Cooksy, 2013). The indicators chosen fluctuated widely from year to year without a clear relationship with performance. The use of the chosen indicators for resource allocation led to unwanted effects on the way that performance was reported. The need to claim successful impacts from agricultural research has taken on a life of its own, often leading to claims that are later contested (Sumberg *et al*., 2012). Thus, although a metrics and indicator system could form part of a performance management system, the timescale for delivering measurable impacts may preclude this being central to the evaluation of CRPs, CGIAR Centers or individual staff.

Bibliometric indicators are often used to evaluate the scientific impact of research articles, individual scientists and research groups or institutes, and this type of analysis has become a special field of research. Citation analyses, impact factors or other indicators are often criticized for not including the analysis of reports and articles outside the domain of peer review, although they may play important roles in influencing policy, etc. Altmetrics[5] can be applied to both documents and data sets if they have their unique digital object identifier (DOI), although there is considerable controversy concerning the use of altmetrics (Colquhoun and Plested, 2014). It would seem logical that CGIAR set up its own DOI management system to support the dissemination of its research and traceability of all of its uses.

Funding agencies may require confirmation that staff time charged to specific projects was duly invested in those projects. The panel noted with dismay the statement in the SRF Management Update (CGIAR Consortium, 2013b) signaling the intention to establish "… a system of time recording … that supports reliable allocation of staff time to outcomes that can be monitored and verified." More than a decade of experience of time-writing systems in different organizations suggests that this rapidly becomes a form of institutionalized fraud. Apart from adding a time-consuming extra layer of bureaucracy that is strongly resisted by scientists, time-writing systems do little more than hand the role of control of staff time to financial auditors. Such systems are time-consuming and there is little evidence that they produce useful information for research management or to improve efficiency or delivery. The opposite is more likely to be true. Given the increasing alignment of priorities of funding agencies around the CRPs, a degree of trust is needed concerning the ways in which scientists invest their time.

5. Altmetrics, or 'alternative metrics,' are an emerging field of new methods for measuring the use and importance of scholarly articles, particularly in the sciences. As opposed to more traditional bibliometrics, such as Impact Factor, altmetrics provide article-level data and are based on new electronic sources of information, such as number of downloads and page views from a publisher, repository or online reference manager like Mendeley, or the amount of discussion generated in online platforms such as Twitter or blogs.

7. CGIAR metrics and indicator systems as an integrating framework

Constructing a robust and coherent indicator framework can be a challenge. It is particularly important to be clear about how metrics are gathered, linked and integrated. An essential part of an indicator framework for CGIAR is to clearly establish the linkages between the CRP research outputs and the IDOs and SLOs. That is, as the system develops, the theories of change supporting the chosen research activities need to evolve into quantitative models. Vagueness in the analytical chain can lead to highly tenuous estimates of impact. It is therefore important to be very explicit about the assumptions that underpin the links between different metrics and indicators. These assumptions are evident at early stages in data collection and can impact all subsequent integrative and presentational tasks. Also important for managing and understanding the results chain is the notion of layering and overlap of information availability. For example, it is important that information is available in spatially disaggregated forms across the results chain so that it becomes possible to trace outcomes in and across geographies and IDO and SLO domains.

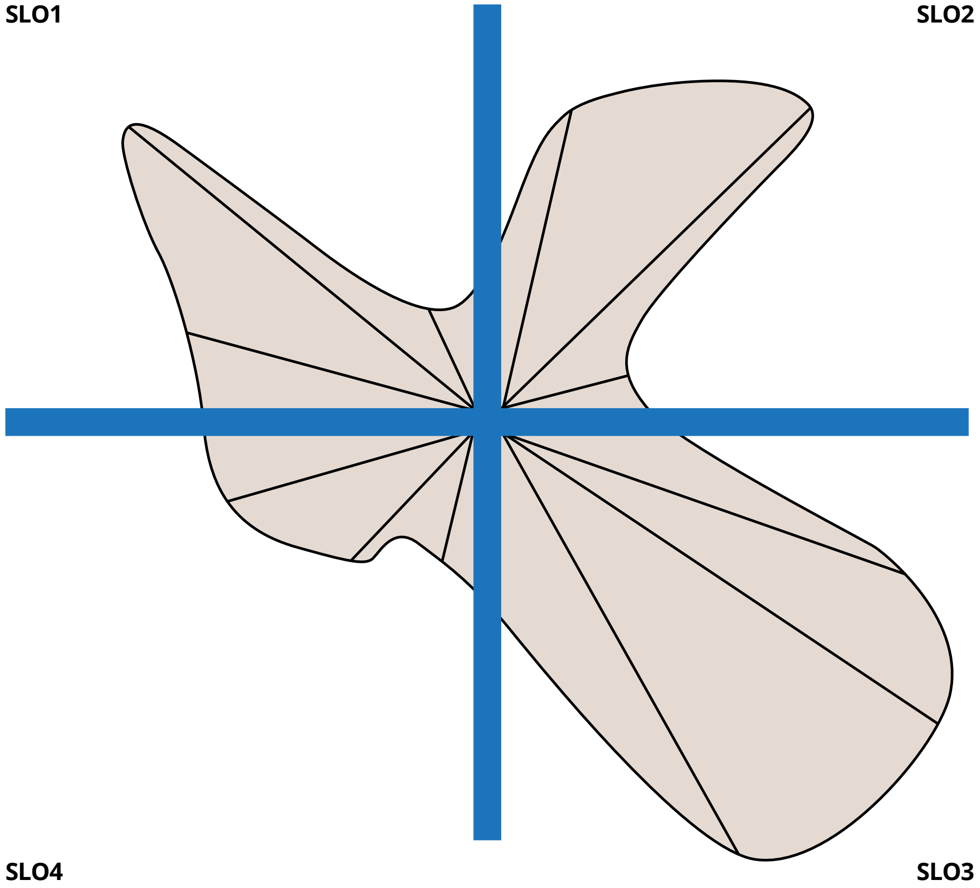
The SLOs provide the basis for a united reporting process: in short, a methodological framework. There are many questions about the organization of the framework, e.g. how much local and how much centralized control should be allowed? Researchers often express a strong desire for comprehensive central databases that have reliable and useable data (e.g. Clapp *et al*., 2013). HarvestChoice has done this to some extent. However, the effective use of such data often requires expert knowledge that is only available 'locally' (e.g. within CRPs). However organized, the value placed on a quantitative system that supports assessment of CGIAR's contribution to the SLOs is that "Even relatively naïve quantitative models outperform expert intuition in a surprising variety of tasks." To transform data into indicators and metrics assumes that the data exist and that a credible model exists to do so.

A useful device for visualizing outputs is a dashboard. Dashboards are well established in the literature (e.g. see Prahalad and Hamel, 1990) and a deceptively simple dashboard is an amoeba diagram (see Figure 7.1). Each 'arm' of the amoeba represents an agreed indicator. The indicators can be compared easily with and between indicators of SLOs. In this manner the amoeba, although requiring extensive preparation and agreement on indicators, provides an easily understood and effective means for comparing indicators against each other. The amoeba can be further nuanced by including a normative or accepted 'band' of equilibrium or sustainability. Such a band provides a target or ideal result for each indicator and this can provide an easy reference to compare with the amoeba. The amoeba is a composite indicator that retains all the explicit details of its components. It has proved effective in a number of contexts (e.g. Bell, 2011; Bell *et al*., 2013; Cassar *et al*., 2013; Coudert *et al*., 2011). The indicators can be assessed in terms of SLO quadrants and could be operationalized at any level. In this sense, a nested hierarchy of interrelated amoebas can be conceived, from SLO to CRP.

It is rare for a monitoring exercise to have consistent and unproblematic access to metrics at all times. The amoeba dashboard methodological process can accommodate breakdowns as circumstances intervene, including changes in agreed metrics (the methodological framework will undoubtedly need periodic refreshing and renewal), changes in data collection processes and other accidental and/or unavoidable changes to the research environment. The overarching concern is to provide a scalable harmonization model that allows most of the indicators to be produced most of the time.

The dashboard depends on harmonization of indicators that are based on coherent theories of change and the formation of small sets of indicators representative of the SLO quadrants that are broadly accepted to

Figure 7.1. The amoeba diagram



Assumes 3 indicators per SLO

represent the goals of CGIAR. Another important requirement is institutional support for learning about agricultural development. For this reason, coherence and learning are seen as a core pragmatic 'set' within the overall metrics system. A governance process is needed for the metrics and indicator system to operate. Such a process would manage issues of urgency, alignment, partnership and governance while being open to the innovation of the overall design.

8. Conclusions and recommendations

To support its mandate of research for development, CGIAR has contributed both concepts and data to numerous national and international information systems on agriculture and natural resources. It has been responsible for establishing many databases on its mandate crops and fields of research. Nevertheless, reviews by the ISPC and feedback from donors and users of data suggest that there is still a major need for improvement in all areas of data, metrics, indicators and information management, access and use. **There are abundant examples of weak quality data, duplication and redundancy in data gathering, poor curation and storage. Obstacles persist that prevent scientists and other users from accessing CGIAR data, metrics and indicators.** Over and above all of this is a widespread weakness in long-term archiving. The move to shorter-term funding cycles and the high degree of mobility of scientists has led us to a situation where locating data from past research is difficult or impossible.

The CGIAR Centers, the Consortium, CRPs and donors all recognize that there are significant areas for improvement in CGIAR management of data, metrics and indicators. The result has been a proliferation of internal and external initiatives to improve the situation. Significant improvements have been made particularly at the level of the CRPs. All the CRPs now have proposals for metrics to monitor their progress toward achieving the IDOs and SLOs. Valuable progress has been made in developing an open access policy and beginning its implementation. The ISPC review has therefore been working at a time of rapid innovation and change. This Panel Report has explored numerous issues that CGIAR will have to address in improving its data, metrics and indicators practices. The major conclusions and recommendations of the panel are as follows.

The present situation. CGIAR has much to be proud of in terms of its contribution to data and metrics on agricultural systems broadly defined. However, there is no system-wide capacity to collect, archive and store data. Access to historical data is difficult and there are major problems with data quality throughout the system. The current emphasis on demonstrating impact has stimulated a plethora of new approaches to measurement and monitoring at all levels. Yet at the same time, the overall purpose of these activities becomes lost in the blur of activity. The panel recommends that all new research programs should be carefully scrutinized to ensure that they have adequate provision for curation, quality control and archiving of data, and for making data, metrics and indicators available to partners and other users. **The panel further recommends that the Consortium provide a normative and control function and ensure periodic peer review of data, metrics and indicators. The Consortium will need to establish approaches to overseeing data, metrics and indicators throughout the system. The provision of comprehensive, easily accessible high-quality data and metrics on agricultural systems should be a major public good; an important product of CGIAR.**

Special challenges of the reformed CGIAR. The portfolio of activities within CGIAR has expanded dramatically in recent years. A broad range of NRM issues is now being addressed. Research is being conducted by very large partnerships composed of diverse types of institutions and users. The SLOs and IDOs are inherently difficult to measure. These changes in CGIAR provide both a need and opportunity for improving the data, metrics and indicators functions. **The panel recommends that the data, metrics and indicators procedures within CGIAR be comprehensive and address all dimensions of research and development activity. A focus on 'systems' (e.g. lowland tropics), which has been adopted by a number of CRPs, creates its own special needs and opportunities for data, metrics and indicators. There will be a need to move beyond the traditional focus on metrics related to yields and varietal adoption.**

Learning and accountability. Data, metrics and indicators are required throughout CGIAR to provide for both learning and accountability. Different metrics are required for each of these major functions. However, all should be included in an integrated data, metrics and indicators system. Considerable confusion has occurred because the term 'metrics' is being used differently by different constituencies. In general, metrics for accountability will be linked to predetermined outcome targets. A learner focus will require the tracking and detection of patterns in metrics accumulated over time. **The panel recommends that the need for both a learner and accountability focus be fully recognized in a data, metrics and indicators system, and that no single objective should dominate.**

Issues of scale. Inevitably CGIAR will need to handle data and metrics at various spatial and temporal scales. Short-term metrics will be required to address issues of accountability. Metrics for learning and research and to measure progress toward the SLOs and IDOs will operate over much longer time frames. The panel would caution against excessive effort being invested in short-term metrics to satisfy needs for accountability. Both for accountability and learning, the development of rigorous long-term data sets is a priority. Data and metrics will also be collected at various spatial scales, ranging from the farm to the planet. Ideally, these metrics should be collected in ways that allow them to be aggregated and disaggregated to respond to questions posed at the different scales. In reality, there are significant scientific challenges in achieving this. **The panel recommends that special attention be given to the problems of aggregation and disaggregation of data collected at different spatial and temporal scales.**

Ontology. It is essential that data, metrics and indicators are properly described and that methods for their collection and curation are carefully documented. The panel recommends that a comprehensive ontology for data and metrics systems in CGIAR be developed by the Consortium with contributions from all CRPs. The ontology should not be developed in isolation from data and metrics work being conducted by other organizations. Although a universal ontology for agricultural data and metrics is some distance in the future, CGIAR could play an important role in moving toward broadly acceptable ontologies. **The panel therefore recommends that the Consortium lead a process to develop a shared ontology.**

The strategy and results framework. The development of a system of data, metrics and indicators is occurring at the same time that the SRF is being refined and that the links between the SRF and the SLOs and IDOs are being made explicit. **The panel considers it important that work on data, metrics and indicators is fully integrated with the process of developing the SRF and the outcome targets.**

SLOs and IDOs. Metrics are clearly required to measure progress toward SLOs and IDOs. However, as these concepts are presently defined, they are not amenable to easy measurement. It will not be easy to attribute changes in the SLOs and IDOs to activities of CGIAR, yet CGIAR needs to be able to measure its contribution to change. **The panel recommends pragmatism in the use of metrics to measure progress toward these outcomes.** At present the metrics debate may be excessively focused on SLOs and IDOs. Considerable care will be needed to make sure that the development of metrics systems for CGIAR is not excessively dominated by the need to measure progress toward the SLOs and IDOs. Accountability metrics should, in the short term, focus on immediate development outcomes and recognize the difficulty of addressing the needs of measuring long-term SLOs.

Impact assessments. CGIAR has a long record of effective impact assessment. It can rightfully claim to be better at impact assessment than many other providers of research and development assistance. Nevertheless, the need for improvements in impact assessment is widely recognized. The new Special Program for Strengthening Impact Assessment in CGIAR (SIAC) is therefore very welcome. SIAC recognizes the need to broaden the range of metrics used to assess impact. In particular, extra attention will be paid in the future to measuring impact on natural resource systems. **The panel strongly endorses the work of SIAC and encourages it to prioritize the issue of evaluating NRM research projects.**

Prioritizing research. A key use of the impact evaluation system is to prioritize research investment. This is driven by the idea that the efficiency of research for development can be increased by investing most in projects that give the 'best bang for the buck' in economic terms. Some donors require a projected return on investment as a key criterion for vetting and selecting projects for funding. This raises a dilemma for priority-setting. When agricultural technologies, such as new crop varieties, are adopted, the poorest people often benefit the least in absolute terms, though they may benefit the most in relative terms – as a proportion of their income or an increase in months of food self-sufficiency. Many development agencies have retreated from trying to work directly with the poorest of the poor due to the intractable and multifaceted nature of the problem. Yet given that SLO1 – reducing rural poverty – is a key goal of CGIAR, does this mean that research targeting the poorest households will not be a priority? **The panel recommends that CGIAR not retreat from working on difficult problems based on arguments couched in simple economic returns.**

Partnerships. Substantial emphasis – correctly in our view – is placed on a wide range of partnerships through which CGIAR will achieve its goals. This means that CGIAR depends on the performance of both research and development partners in achieving its impact. Does a new prioritization around impact mean that CGIAR will avoid weaker partners, such as the national agricultural research and extension systems, which often suffer from chronic underfunding? Building the capacity of such partners could be critical for the long-term sustainability of research outcomes. **The panel welcomes the forthcoming review of partnership arrangements planned by the ISPC.**

Alignment with other initiatives. The development of a comprehensive system of data, metrics and indicators is occurring at a time of numerous other initiatives with similar or overlapping objectives. It is important that CGIAR takes note of these initiatives and, to the extent that it is appropriate, aligns its own work with them. **The panel particularly recommends that CGIAR monitor progress in the development of the United Nations Sustainable Development Goals (SDGs) and their associated targets and indicators.** The panel doubts that the indicators adopted at the political and international level will directly serve the purposes of CGIAR but they will provide context and guidance for the work of CGIAR. As a minimum, the CGIAR data, metrics and indicators should align with the SDGs.

The open access policy. The panel strongly approves of the intentions behind the new open access policy developed by the CGIAR Consortium. This initiative is long overdue and will provide a valuable public good product based on CGIAR research. The panel considers that a degree of pragmatism will be required in the early stages of implementing the open access policy. One feature of the open access system will be to expose CGIAR data and metrics to external scrutiny. This will place pressure on research teams to ensure that their data are of the highest quality, their ontology is correct and that the curation and archiving of the data is of a high standard. **The panel recommends that the main structure of the open access arrangements be put in place rapidly, but since change is so rapid in this domain, the panel notes that it will be important to build in flexibility and the ability to adapt to changing circumstances as the policy implementation proceeds.**

Building a CGIAR resource for the future. Our review reveals the huge potential that a comprehensive open CGIAR database can provide. Ensuring the establishment and use of a coordinated system has to be of greatest priority for CGIAR. Key to the success of such an initiative is that basic, building-block data are made available to allow new approaches to derivation of metrics and indicators in the future. **It is**

essential to measure the basics, and to measure them well. Where more complex metrics and indicators are presented, we propose two principles: (i) All input data should be standardized and made available, including details of the methods used for data collection; and (ii) The calculations used to derive metrics and indicators should be presented transparently. We look forward to CGIAR assuming its leading role in building a comprehensive open database, which will be a critical legacy of CGIAR research for the future.

References

(Including sources cited in annexes)

Alston, J.M., Andersen, M.A., James, J.S. & Pardey, P.G. 2010. *Persistence pays: U.S. agricultural productivity growth and the benefits from public R&D spending*. New York, NY, Springer.

Bell, S. 2000. Logical frameworks, Aristotle and soft systems: A note on the origins, values and uses of logical frameworks, in reply to Gasper. *Public Admin. Develop*., 20: 29–31.

Bell, S. 2011. From sustainable community to big society: Ten years learning with the Imagine Approach. *IRGEE*, 20(3): 247–267 (available at: http://dx.doi.org/10.1080/10382046.2011.588509).

Bell, S., Correa Pena, A. & Prem, M. 2013. Imagine coastal sustainability. *Ocean Coast Manag*., 83: 39–51.

Bergek, A., Jacobsson, S., Hekert, M. & Smith, K. 2010. Functionality of innovation systems as a rationale for and guide to innovation policy. *In* Smits, R. e*t al. The theory and practice of innovation policy – An international research handbook*. London, Edgar Elgar.

Carletto, C., Gourlay, S. & Winters, P. 2013a. *From guesstimates to GPStimates: Land area measurement and implications for agricultural analysis*. World Bank Policy Research Working Paper 6550. The World Bank Development Research Group, Poverty and Inequality Team (available at www.worldbank.org/en/research).

Carletto, C., Savastano, S. & Zezza, A. 2013b. Fact or artifact: The impact of measurement errors on the farm size–productivity relationship. *J. Dev. Econ*., 103: 254–261.

Cash, D.W., Adger, W.N., Berkes, F., Garden, P., Lebel, L., Olsson, P., Pritchard, L. & Young, O. 2006. Scale and cross-scale dynamics: Governance and information in a multilevel world. *Ecol. Soc*., 11(2): 8 (available at www.ecologyandsociety.org/vol11/iss11/art18/).

Cassar, L.F., Conrad, E., Bell, S. & Morse, S. 2013. Assessing the use and influence of sustainability indicators at the European periphery. *Ecol. Indic*., 35: 52–61.

CGIAR. 2010. A Strategy and Results Framework for the CGIAR (available at www.cgiar.org/www-archive/www.cgiar.org/changemanagement/pdf/cgiar_srf_june7_2010.pdf).

CGIAR Consortium. 2013a. Open Access and Data Management Policy. Montpellier (available at http://library.cgiar.org/bitstream/handle/10947/2875/CGIAR%20OA%20Policy%20-%20October%202%202013%20-%20Approved%20by%20Consortium%20Board.pdf).

CGIAR Consortium. 2013b. The CGIAR Strategy and Results Framework Management Update, 2013–2014. Montpellier (available at http://library.cgiar.org/bitstream/handle/10947/2895/2014%20-%20SRF%20Management%20Update_Final_2013_12_20.pdf).

Clapp, A., DauSchmidt, N., Millar, M., Hubbard, D. & Shepherd, K. 2013. *A survey and analysis of the data requirements for stakeholders in African agriculture*. Nairobi, World Agroforestry Centre (ICRAF).

Colquhoun, D. & Plested, A.J. 2014. Why you should ignore altmetrics and other bibliometric nightmares. DC's Improbable Science blog (available at www.dcscience.net/?p=6369).

Coudert, E. & Larid, M. 2011. *IMAGINE: A set of tools and methods to assist integrated coastal zone management in the Mediterranean*. Sophia Antipolis, France, Blue Plan UNEP/MAP Regional Activity Centre.

de Janvry, A. & Sadoulet, E. 2002. World poverty and the role of agricultural technology: Direct and indirect effects. *J. Dev. Stud.*, 38: 1–26.

EU SCAR. 2012. *Agricultural knowledge and innovation systems in transition – A reflection paper*. Brussels, Standing Committee on Agricultural Research (SCAR) Collaborative Working Group on Agricultural Innovation and Knowledge Systems (CWG-AKIS).

Ewert, F., van Ittersum, M., Heckelei, T., Therond, O., Bezlepkina, I. & Andersen, E. 2011. Scale changes and model linking methods for integrated assessment of agri-environmental systems. *Agric. Ecosyst. Environ.*, 142: 6–17.

Foster, J., Greer, J. & Thorbecke, E. 1984. A class of decomposable poverty measures. *Econometrica*, 52: 761–766.

Fuglie, K.O. 2012. Productivity growth and technology capital in the global agricultural economy. *In* K.O. Fuglie, S.L. Wang & V.E. Ball, eds. *Productivity growth in agriculture: An international perspective*. Wallingford, UK, CAB International.

Gassner, A., Alvare, L.M., Bamba, Z., Beare, D., Bernardo, M., Biradar, C., van Brakel, M., Chapman, R., Dileepkumar, G., Dieng, I., Erlita, S., Fulss, R., Poole, J., Kshatriya, M., Selim, G., Simon, R., Sonder, K., Prasa, N., Garruccio, M., Staiger Rivas, S., Rajasekharan, M. & Rao, C.S. 2010. *Shifting the goalposts – From high impact journals to high impact data*. White paper. CGIAR Consortium (available at www.worldagroforestry.org/downloads/publications/PDFs/PP13223.PDF).

Hansen, M.C, Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O. & Townshend, J.R.G. 2013. High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160): 850–853 (available at www.sciencemag.org/content/342/6160/850).

Hazell, P. & Haddad, L. 2001. *Agricultural research and poverty reduction: Food, agriculture, and the environment*. Discussion Paper 34. Washington, DC, IFPRI, TAC-CGIAR.

Hillenbrand, E. 2010. Transforming gender in homestead food production. *Gender & Development*, 18(3): 411–425.

Hubbard, D.W. 2010. *How to measure anything: Finding the value of 'intangibles' in business*. 2nd edition. Hoboken, NJ, John Wiley & Sons.

Immonen, S. & Cooksy, L.L. 2013. Using performance measurement to assess research: Lessons learned from the international agricultural research centres. *Evaluation*, 20: 96–114

ISEAL Alliance. 2012. *Growth of the sustainability standards movement*. issuu online (available at http://issuu.com/isealalliance/docs/growth_of_the_sustainability_standards_movement_ti/1?e=0).

ISPC. 2009. Stripe review of the social sciences in the CGIAR. Rome, CGIAR Independent Science and Partnership Council Secretariat (available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1844803).

ISPC. 2011. CGIAR Science Forum 2011: The Agriculture–Environment Nexus: Summary. Beijing, 17–19 October. Rome, CGIAR Independent Science and Partnership Council Secretariat (available at www.sciencecouncil.cgiar.org/system/files_force/ISPC_MobilizeScience_SF2011_Summary.pdf?download=1).

ISPC. 2012a. A Stripe review of natural resources management research in the CGIAR. Rome, CGIAR Independent Science and Partnership Council Secretariat (available at: www.sciencecouncil.cgiar.org/system/files_force/ISPC_StrategyTrends_NRM_StripeReview_0.pdf).

ISPC. 2012b. *Strengthening Strategy and Results Framework through prioritization.* White paper. Rome, CGIAR Independent Science and Partnership Council Secretariat (available at www.sciencecouncil.cgiar.org/system/files_force/ISPC_WhitePaper_Prioritization.pdf).

ISPC. 2012c. *Strategic overview of CGIAR Research Programs*. Part I. *Theories of change and impact pathways*. Rome, CGIAR Independent Science and Partnership Council Secretariat (available at: www.sciencecouncil.cgiar.org/system/files_force/ISPC_WhitePaper_TOCsIPs.pdf.

ISPC. 2013a. Concept Note for a strategic study on metrics, benchmarking, and monitoring for CRP evaluation and impact assessment. Rome, CGIAR Independent Science and Partnership Council Secretariat (available at www.sciencecouncil.cgiar.org/system/files_force/ISPC_StrategyTrends_Metrics_ConceptNote.pdf).

ISPC. 2013b. *CGIAR System-Level Outcomes (SLOs), their impact pathways and inter-linkages.* White paper. Rome, CGIAR Independent Science and Partnership Council Secretariat (available at: www.sciencecouncil.cgiar.org/system/files_force/ISPC_WhitePaper_SLOsIPs.pdf).

ISPC. 2013c. *CGIAR Science Forum 2013: Human nutrition and health: Targets for agricultural research: Summary*. Bonn, 23–25 September. Rome, CGIAR Independent Science and Partnership Council Secretariat (available at http://scienceforum13.org/sites/default/files/docs/SF13-summary.pdf).

ISPC. 2013d. ISPC comments on the CGIAR Open Access Policy (June 2013 draft for review). Rome, CGIAR Independent Science and Partnership Council Secretariat (available at www.sciencecouncil.cgiar.org/publications).

ISPC. 2013e. *Workshop report: ISPC strategic study on metrics, benchmarking, and monitoring for CRP evaluation and impact assessment: A brainstorming stakeholder workshop*. FAO, Rome, 10–11 December 2013. Rome, CGIAR Independent Science and Partnership Council Secretariat (available at https://sites.google.com/site/metricsmeeting/).

Janssen, S., Athanasiadis, I.N., Bezlepkina, I., Knapen, R., Li, H., Domínguez, I.P., Rizzoli, A.E. & van Ittersum, M.K. 2011. Linking models for assessing agricultural land use change. *Comput. Electron. Agric.*, 76: 148–160.

Janssen, S., Andersen, E., Athanasiadis, I.N. & van Ittersum, M.K. 2009. A database for integrated assessment of European agricultural systems. *Environ. Sci. Policy*, 12: 573–587.

Jerven, M. 2013. *Poor numbers: How we are misled by African development statistics and what to do about it*. Ithaca, NY, Cornell University Press.

Kilic, T., Zezza, A., Carletto, C. & Savastano, S. 2013. *Missing(ness) in action: Selectivity bias in GPS-based land area measurements*. World Bank Policy Research Working Paper 6490. The World Bank Development Research Group, Poverty and Inequality Team (available at www.worldbank.org/en/research).

Letouzé, E. 2012. *Big data for development: Challenges and opportunities*. New York, NY, UN Global Pulse (available at www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf).

Markevičius, A., Katinas, V., Perednis, E. & Tamašauskienė, M. 2010. Trends and sustainability criteria of the production and use of liquid biofuels. *Renew. Sust. Energ. Rev.*, 14: 3226–3231.

Masters, W.A. 2013. *Synthesis of Foresight Study on Urbanization and farm size in developing countries: Implications for agricultural research*. Paper commissioned by the CGIAR ISPC (available at http://sites.tufts.edu/willmasters/files/2012/09/WillMasters_ISPC-FarmSizeForesightWorkshopSynthesis_12May2013.pdf).

Mock, N., Papendieck, A. & Morrow, N. 2013. From complexity to food security: Decision support: Novel methods of assessment and their role in enhancing the timeliness and relevance of food and nutrition security information. *Global Food Secur.*, 2: 41–49.

Morrow, N., Mock, N. & Papendieck, A. 2011. *Ushahidi Haiti project evaluation final report*. Ushahidi [website] (available at http://blog.ushahidi.com/2011/04/19/ushahidi-haiti-project-evaluation-final-report/).

OECD. 2012. *Agricultural innovation systems: A framework for analyzing the role of the government*. Paris, Organisation for Economic Co-operation and Development.

OECD & Eurostat. 2005. *Oslo manual: Guidelines for collecting and interpreting innovation data*. OECD Publishing (available at www.oecd.org/innovation).

Prahalad, C.K. & Hamel, G. 1990. The core competence of the corporation. *Harvard Bus. Review*, 68(3): 79–91.

RSB. 2011. *Indicators of compliance for the RSB EURED Principles and Criteria* [RSB-IND-11-001-20-001 (Version 2.0)]. Lausanne, Roundtable on Sustainable Biofuels (available at http://rsb.org/pdfs/standards/RSB-EU-RED-Standards/10-12-13-RSB-IND-11-001-20-001-RSB-EU-RED-Indicators-vers-2-0.pdf).

RSPO. 2013. *Principles and criteria for the production of sustainable palm oil Including indicators and guidance*. Roundtable on Sustainable Palm Oil (available at www.rspo.org/file/PnC_RSPO_Rev1.pdf).

Scobie, G.M. & Rafael Posada, T. 1977. The impact of technical change on income distribution: The case of rice in Colombia. *Am. J. Agric. Econ*., 60: 85–92.

SDSN. 2013. *An action agenda for sustainable development. Report to the UN Secretary-General*. New York, NY, Sustainable Development Solutions Network (available at http://unsdsn.org/files/2013/11/An-Action-Agenda-for-Sustainable-Development.pdf).

SDSN. [2014]. *SDSN TG7 Issue Brief: Goals, targets, and indicators for sustainable agriculture*. (Prepared by the Thematic Group 7 Sustainable Agriculture and Food Systems) New York, NY, Sustainable Development Solutions Network (available at http://unsdsn.org/wp-content/uploads/2014/02/Issue-Brief-Goals-Targets-and-Indicators-for-Sustainable-Agriculture.pdf).

Shepherd, K., Farrow, A., Ringler, C., Gassner, A. & Jarvis, D. 2013. *Review of the evidence on indicators, metrics and monitoring systems*. Nairobi, World Agroforestry Centre.

Sumberg, J., Irving, R., Adams, E. & Thompson, J. 2012. Success making and success stories: Agronomic research in the spotlight. *In* J. Sumberg & J. Thompson, eds. *Contested agronomy: Agricultural research in a changing world*, pp. 186–203. London, Earthscan.

van Dam, J., Junginger, M., Faaij, A., Jurgens, I., Best, G. & Fritsche, U. 2006. Overview of recent developments in sustainable biomass certification. *IEA Bionenergy Task* 40.

van Galen, M.A. & Poppe, K. 2013. Innovation monitoring in the agri-food business is in its infancy. *EuroChoices*, 12(1): 28–29.

van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P. & Hochman, Z. 2013. Yield gap analysis with local to global relevance – A review. *Field Crop. Res*., 143: 4–17.

van Ittersum, M., Ewert, F., Heckelei, T., Wery, J., Alkan Olsson, J., Andersen, E., Bezlepkina, I., Brouwer, F., Donatelli, M., Flichman, G., Olsson, L., Rizzoli, A., van der Wal, T., Wien, J.E. & Wolf, J. 2008. Integrated assessment of agricultural systems: A component-based framework for the European Union. *Agric. Syst*., 96, 150–165.

van Noordwijk, M., Agus, F., Dewi, S. & Purnomo, H. 2014. Reducing emissions from land use in Indonesia: Motivation, policy instruments and expected funding streams. *Mitig. and Adapt. Strateg. for Global Change*, 19(6): 677–692 (available at http://link.springer.com/article/10.1007%2Fs11027-013-9502-y).

Volk, M. & Ewert, F. 2011. Scaling methods in integrated assessment of agricultural systems: State-of-the-art and future directions. *Agric. Ecosyst. Environ*., 142: 1–5.

Walker, T.S. 2000. Reasonable expectations on the prospects for documenting the impact of agricultural research on poverty in *ex-post* case studies. *Food Policy*, 25: 515–530.

White, J.W. & van Evert, F.K. 2008. Publishing agronomic data. *Agron. J.*, 100: 1396–1400.

Zamir, D. 2013. Where have all the crop phenotypes gone? *PLoS Biol*., 11(6): e1001595 (doi:10.1371/journal.pbio.1001595).

Zarilli, S. 2008. Overview of certification developments and comparison of schemes. In *Making certification work for sustainable development: The case of biofuels*, pp. 1–20. United Nations Conference on Trade and Development. New York and Geneva, United Nations (available at http://unctad.org/en/Docs/ditcted20081_en.pdf).

Annexes

Annex 1. Glossary of terms and definitions

Altmetrics	New metrics proposed as an alternative to the widely used journal impact factor and personal citation indices like the h-index. Although altmetrics are often thought of as metrics about articles, they can be applied to people, journals, books, data sets, presentations, videos, source code repositories, web pages, etc. Altmetrics cover not just citation counts, but also other aspects of the impact of a work, such as how many data and knowledge bases refer to it, article views, downloads, or mentions in social media and news media.
Attribution	The extent to which observed (or expected to be observed) changes can be linked to a specific intervention after controlling for the effects of other interventions or confounding factors.
Baseline survey/ study	An analysis describing the situation in a project area – including data on individual primary stakeholders – prior to a research and development intervention. Comparisons can later be made between new data and baseline data, to monitor project progress.
Benchmark	Reference point or standard against which performance or achievements can be compared; it might refer to what has been achieved in the past, or what could reasonably have been achieved under the circumstances.
Benchmark sites	Selected research-for-development sites that are large enough to capture typical variation in agroecological and socioeconomic conditions found in the wider agroecological zone. (e.g. CRP Drylands and Humidtropics)
Data	Raw data are observations, such as weight, height, plot size.
Hub	Local innovation system (involving researchers, farmers, agro-enterprises, extension) that accelerates adaptation, testing and scale-out of agro-technology and research-to-farmer communication approaches, which are adapted to resource-poor farmer environments. (e.g. CRP Wheat & Maize)
Impact	The ultimate planned and unplanned consequences of a program; an expression of the changes actually produced as a result of the program, typically several years after the program has stabilized or been completed. Impact can be positive or negative, intended or unintended.
Impact assessment	The process of assessing the impact of a program in an intervention area.
Indicator	Quantitative or qualitative factor or variable that provides a simple and reliable basis for assessing achievement, change or performance. A unit of information that can be measured at different points in time to help show changes in a specific condition. Progress toward a given goal or objective can be monitored using multiple indicators. See Section 1.1.
Information management system	A system of inputting, collating and organizing data that should provide selected data and reports for review by the project managers, to assist in monitoring and controlling project planning, resources, activities and results.

Information value stream	The set of activities linking the information project to its ultimate use, such that it provides support to decision-making (see Mock *et al.*, 2013).
Innovation	The creation of better or more effective products, processes, services, technologies or ideas (see Box 4.1 for more details).
Innovation system	This concept stresses that the flow of technology and information among people, enterprises and institutions is key to an innovative process. It contains the interaction between the actors who are needed in order to turn an idea into a process, product or service available on the market (see Box 4.1).
Intervention	Any promotive, preventive, curative or rehabilitative activity where the primary intent is to improve conditions.
Measurement	The assignment of numbers to objects or events; all measurements consist of three parts: magnitude, dimensions (units) and uncertainty.
Metric	Metrics are computed by aggregating and combining raw data, for example, yield or height-for-age. They often represent the values on which indicators are built.
Monitoring	The regular collection and analysis of information about a program, project or activity to assist timely decision-making, ensure accountability and provide the basis for evaluation and learning.
Monitoring and evaluation	The combination of monitoring and evaluation which together provide the knowledge required for: (a) effective project management, and (b) reporting and accountability responsibilities.
Outcome	An effect or consequence of a program in the medium term, often considered to be the result 5 years or more after the start of the program or intervention. A medium-term result that is the logical consequence of achieving a combination of short-term outputs.
Output	The tangible, short-term and intended result to be produced through sound management of the agreed inputs; also includes changes resulting from the research intervention that are needed to achieve the intended outcome later.
Performance	The degree to which an R&D intervention or a partner operates according to specific criteria/standards/guidelines or achieves results in accordance with stated goals or plans.
Risk	Possible negative external factors, i.e. events, conditions or decisions, which are identified as having the potential to seriously delay or prevent the achievement of project objectives and outputs (and which are normally largely or completely beyond the control of the project management).
Sentinel landscapes	A site or a network of sites, geographically or issue bounded, where a broad range of biophysical, social, economic and political data are collected with consistent methods and are monitored and interpreted over the long term. (e.g. CRP FTA)
Trade-off	An exchange of one thing in return for another; especially relinquishment of one benefit or advantage for another regarded as more desirable.
Uncertainty	The lack of complete certainty or the existence of more than one possibility. The 'true' outcome/state/result/value is not known (Hubbard, 2010).

Annex 2. Summary of findings of a questionnaire survey conducted in October 2013 on CRP metrics

CRP	Site characterization; Baseline data collection	Methods of data collection	Major gaps; Challenges	Coordination with other CRPs; Opportunities for data sharing	Data collection cost & responsibility sharing – Ownership & Attribution	Partners outside CGIAR
1.1. Drylands	Basic site characterization done for Action sites (2012). Indicators being developed (i.e. for land degradation, sustainable intensification). Baseline data collection for households (HH) ongoing.	Numerous methods (on-station and on-farm trials, HH surveys, remote sensing and use of existing databases). Indicators to measure progress on research outputs will include databases, research reports, communications, capacity building and peer-reviewed publications.	System CRPs are on uncharted ground; need time to develop metrics.	Main efforts are with other system CRPs; gradual development of research plans in areas where CRPs coincide. Main ongoing effort is in Burkina Faso and includes WLE, FTA, Grain Legumes and Dryland Cereals.	Data will be shared through the Statistical Services Centre (SSC) at Reading University. Frameworks (e.g. systems models) through which data can be shared but relations have to be further developed.	SSC at Reading University
1.2. Humidtropics	Characterization of Action sites ongoing as sites are being developed. Some indicators identified for characterization and baselines, with emphasis on existing data and knowledge sources. Developed new approaches and started implementing them in 3–4 Action areas.	Most data from existing studies such as the Consortium for Improving Agriculture-based Livelihoods in Central Africa and Sub-Saharan Africa (SSA) Challenge Programme as these programs have been absorbed into Humidtropics. Methods include: • review of existing data sets from earlier projects in Action areas; • review of data from other sources; • commissioning of specific data in field sites (including HH surveys; field data gathering; on-farm experimentation). Key metrics being collected are on areas such as productivity, NRM, nutrition and gender roles.	Metrics need to cover trade-offs and interactions across components in systems, e.g. TFP. Livelihood indicators and contribution of off-farm and on-farm activities toward income. Social indicators go beyond those listed and need more cultural variables. Missing variables related to socio-technical regime changes and capacity to innovate at farm, socio-technical regime and landscape levels.	Submitted proposal to pilot Results-Based Management System and agreed with RTB on joint work in a specific Action site (joint workshop, topics including metrics and RBM). With other systems CRPs, scheduled workshop on capacity to innovate (including indicators), March 2014. Registered with Dataverse but yet to deliver data. An online repository and share site (using Alfresco Community Ed., Open Source, UK) where all important documents, data, images, etc. are deposited, annotated and tagged for future access and use.	Expect other CRPs to collect data relevant to their work. If this happens in Humidtropics we would need access to see how they relate to and influence the system (e.g. on TFP and linked trade-offs). We will make such data available so it will improve their decision-making in relation to targeted interventions. Will follow CGIAR Intellectual property guidelines. CRP not being a legal entity, IP issues will refer to regulations within CGIAR and Lead Centre.	No such partnerships yet in data usage, etc.
1.3. Aquatic Agricultural Systems (AAS)	5 flagship sites selected; characterization and prioritization of research for development (R4D) agendas via participatory diagnostic and design processes. Strategic research plans are established in 3 of the 5 sites. Selected baseline measurements made (biodiversity in farming systems, gender baseline) at successive sites. The IDO indicator baseline survey in planning phase for implementation in 2014.	Using existing data (mostly HH surveys, intra-HH data), supplemented by surveys with similar questionnaires to cover intervention sites and special topics not covered by existing information. Examples include: gender values and norms, with information collected via structured questionnaires and focus group discussions (FGDs; community level information) and agricultural biodiversity information collected via structured HH questionnaires and FGDs for community information. The IDO Baseline Survey (planned) will consist of HH surveys and FGDs.	Metrics for community and individual empowerment; innovative capacity.	No coordination activities yet. Data sharing: extensive opportunities at sites where CRPs have common geographic areas of work, e.g. in southwest Bangladesh where various CRPs are engaged.	Cost and responsibility not yet discussed. Will follow CGIAR open access protocols and policy.	USAID Feed the Future surveys (where applicable) World Bank-Country LSMS surveys (where applicable)
2. Policy, Institutions & Markets (PIM)	Indicators being defined.	Mixed methods (see IDO paper for description of draft indicators).	Too early to tell.	Once a common set of indicators is shared, we can explore coordination. At present IDOs have been shared, but not indicators. Data sharing: many opportunities; difficult to respond in detail.	In shared sites, will seek agreement with other CRPs. Follow open data policy.	UN Statistical Office, SDGs, FAOStat, LSMS, national sources, etc. Regular use of these data, partnership with units developing them.

CRP	Site characterization; Baseline data collection	Methods of data collection	Major gaps; Challenges	Coordination with other CRPs; Opportunities for data sharing	Data collection cost & responsibility sharing – Ownership & Attribution	Partners outside CGIAR
3.1. WHEAT	CRP works across developing world, with emphasis on wheat-growing areas in Mexico, Africa and Asia. In selected key countries characterization and baseline data collection has advanced, through HH surveys and sector characterization (through key informants and secondary data). [Details and list of countries given]. • Biophysical characterization: mega-environments/ agroecologies (global). • An overarching baseline at global level is being planned for selected metrics – including variety releases, seed volumes, and estimated adoption – building on past global impact studies by CIMMYT.	On-station trials: CIMMYT International Wheat Nurseries & International Wheat Improvement Network (IWIN), multi-location testing, productivity: Long-term agronomic trials (productivity; environment, including nutrient use efficiency, water-use efficiency; greenhouse gas emissions (south Asia, Mexico). GIS/remote sensing: biophysical characterization; socioeconomic characterization (environment, productivity, livelihood, policy) – e.g. http://wheatatlas.org/. On-farm trials: on-farm agronomic trials; participatory varietal selection (productivity, environment, social [e.g. gender-disaggregation participants]). Monitoring surveys: Global Wheat Rust. Monitoring System (productivity, environment). HH surveys: adoption studies, baseline surveys, panel surveys (productivity, livelihood, social [incl. gender-disaggregation]). Community surveys/FGDs: characterization, monitoring (social, livelihood, productivity, policy). Key informants and secondary data: sector characterization; value chains/ markets; variety release; adoption studies (policy; livelihood; productivity; social). Meta-analysis: particular innovation (e.g. conservation agriculture); Wheat Pedigree (GRIS), Wheat Seed Catalog (productivity).	Geographic coverage: limited geographic coverage with significant data gaps; representativeness of available data [e.g. for secondary data]. Thematic coverage: significant data gaps; studies/methods differ in terms of metrics covered, detail and reliability; linking difficulties between metrics and studies/methods; biophysical vs. socioeconomic indicators; gender (e.g. limited evidence base in wheat-based systems/livelihoods). Metrics characteristics: measurability; scalability; dynamics; reliability. Process coverage: impact pathways and IDOs of research for development; relevant and cost-effective indicators for M&E; minimum data set; attribution; measurement approaches; local to global scale.	MAIZE: close coordination on metrics, joint design of monitoring plans, shared methodologies. CCAFS: shared outputs-to-outcomes methodology; joint methodology on adaptation & mitigation; open access data sharing; collaboration on common sites, data standards, tools, collection and sharing (Agtrials, Dataverse). PIM: Collaboration on data standards, tools, collection & sharing. Across CRP-level: • IDOs Working Group defined set of generic IDOs; Working Group 1 on standardized annual CRP progress reporting including progress indicators, CRP portfolio reporting • Gender: Active engagement in current planning process for cross-CRP collaboration on 'Gender Norms and Agency in Agriculture' [Objectives included]. Wherever significant geographic and thematic complementarities exist, joint coordination and/or implementation (e.g. CCAFS, PIM, MAIZE, system CRPs); joint study on 'Gender Norms and Agency in Agriculture' (across CRP). Agreement and standardization of common metrics and methods. Modular approaches, i.e. ability to add specific modules to complement standard modules with common metrics. Economies of scale in data collection. Thematic complementarities can include biophysical characterization and analysis; socioeconomic characterization and analysis (value chains, policies, adoption and impacts, livelihoods and vulnerability), modeling, system analysis.	Sharing of costs and responsibilities may be feasible wherever significant geographic and thematic complementarities exist and there is willingness among CRPs to agree on modalities. This includes joint funding between CRPs – such as joint positions (e.g. a monitoring, learning and evaluation expert and a gender focal point for MAIZE–WHEAT); joint studies (e.g. Global Futures with PIM); joint study areas/locations (e.g. CCAFS; possibly system CRPs); and joint resource mobilization (e.g. initial discussions with PIM). Data ownership and attribution will follow standard approaches in place between R4D partners – including joint ownership, joint publications wherever feasible/relevant, and at least full acknowledgment. Most R4D partners – including CIMMYT as host center of the CRP – are also moving to open access whereby data will be made publicly available after a lapse of time.	WHEAT partners with various non-CGIAR R4D partners and uses data sources from external sources. Intensive collaboration with NARES programs in all target countries, and to varying degrees, collaboration with the private sector, other research institutes and development partners/ stakeholders (including NGOs). This includes the management and facilitation of data collection and sharing networks. WHEAT variously uses secondary data sources available in the public domain with due acknowledgment. In case of LSMS there are ongoing discussions to try to incorporate information on varietal use and intensification practices – currently lacking or underspecified in the instrument but which would enhance the utility for the CRP.

CRP	Site characterization; Baseline data collection	Methods of data collection	Major gaps; Challenges	Coordination with other CRPs; Opportunities for data sharing	Data collection cost & responsibility sharing – Ownership & Attribution	Partners outside CGIAR
3.2. MAIZE	CRP works across the developing world – with emphasis on maize-growing areas in Mexico, SSA and (sub-) tropical Asia. In selected key countries characterization and baseline collection has advanced, through HH surveys and sector characterization (key informants and secondary data) [Details and list of countries given]. – Biophysical characterization: mega-environments/agroecologies (global).	On-station trials: International Maize Trials Network, Multi-Location Trials, productivity; long-term (LT) agronomic trials (productivity; environment – including nutrient use efficiency, water-use efficiency; greenhouse gas emissions – Mexico). • GIS/remote sensing: as above for WHEAT. • On-farm trials: as above for WHEAT. • Monitoring surveys: e.g. Maize Lethal Necrosis in eastern Africa (productivity; environment). • Household and community surveys/ FGDs: as above for WHEAT. • Meta-analysis: particular innovation (e.g. conservation agriculture).	As above for WHEAT.	WHEAT: Close coordination on metrics, joint design of monitoring plans, shared methodologies. Less prospects for common sites given different geographies/ecologies. CCAFS: As above for WHEAT. PIM: Collaboration on data standards, tools, collection & sharing Across CRP-level: As above for WHEAT. Data sharing: As above for WHEAT.	As above for WHEAT.	As above for WHEAT.
3.3. GRiSP	Baseline surveys and environmental characterizations finished for some hubs, in progress for others. In Asia, a large set of baseline data exists (HH data, environmental characterization) developed from various projects. IRRI is in the process of identifying 'benchmark' sites across the region where metrics will be collected on a systematic basis.	HH surveys, village surveys, on-farm trials, on-station trials, multi-location breeding trials, participatory varietal selection, consumer surveys, value-chain actor surveys (millers, traders, processors, retailers, input suppliers, etc.), remote sensing, GIS, existing statistics (e.g. national statistics, FAO and USDA data on production, import, export, price, yield, etc.), existing information with partners. [Attached a few maps showing some key locations (hubs, experimental platforms, regions, etc.) where data are being collected.]	Major challenge is to aggregate metrics from a project level up to the program level. GRiSP encompasses over 200 bilaterally funded projects, each with its own impact pathway and theory of change, indicators, metrics and donor-reporting requirements. The main challenge is to integrate/ aggregate individual project metrics into an overarching framework for the whole of GRiSP (which operates at a global scale).	Common sites and monitoring and evaluation (M&E) systems through joint projects that cut across CRPs, such as The Cereal Systems Initiative for South Asia (GRiSP/Maize/Wheat), Global Futures (PIM/GRiSP), Challenge Program on Water & Food (GRiSP/AAS/WLE). Data sharing: At joint sites, e.g. in India, Bangladesh, Myanmar, with WLE, AAS, WHEAT and MAIZE specifically.	No concrete plans developed yet. GRiSP encompasses over 900 partners. Co-ownership of data is key to collaboration. A key guiding principle for CGIAR will be the Open Access policy and implementation documents (with some concern expressed about implementation of the policy).	More than 900 R&D partners; many of them involved in data collection. Coordination through our networks, projects, consortia, platforms. GRiSP uses UN agencies such as FAO, UN statistical office as data source where useful.
3.4. Roots, Tubers and Bananas (RTB)	RTB does not have its own sentinel sites. We will use those of systems CRPs. We are in discussion with Humidtropics but have not yet defined indicators or planned baseline data collection. We plan to do this as part of a pilot of results-based management in 2014.	There are multiple HH surveys and on-farm trials among RTB partners. We are developing shared databases (e.g. Agtrials) but so far no consensus on key metrics. Clearly the indicators for IDOs will provide guideline. [Impossible to complete this question because of this diversity. However, as part of adoption studies we are collecting information about areas under specific technologies, losses and yields.]	We do not have a framework for this derived from IDOs. So can't really speak of a gap. But what we urgently need is a cross-partner platform. Would be good if there could be a shared data platform, or at least guidelines for compatibility across CRPs.	Beginning conversation with Humidtropics. Discussed shared attribution and IDOs with A4NH but did not get to metrics. Data sharing: Humidtropics, baselines and adoption; A4NH changes through adoption of micronutrient-rich crops.	Not discussed yet. Ownership and attribution: not discussed yet, although thinking about this in context of IP policy.	Not yet clear.

CRP	Site characterization; Baseline data collection	Methods of data collection	Major gaps; Challenges	Coordination with other CRPs; Opportunities for data sharing	Data collection cost & responsibility sharing – Ownership & Attribution	Partners outside CGIAR
3.5. Legumes	Baseline data and site information presented in the appendices to approved project proposal [details of appendices given]. Baseline data were collected under the Tropical Legumes II (TL-II) project in Ethiopia and Kenya. Other baseline data collection has been completed in Uganda, Rwanda and is ongoing in Ethiopia, Tanzania, Malawi and Zambia. Baseline study on consumption planned for 2014. Gender data are currently being collected.	Baseline data from publicly available data sets, mostly FAO and from HH and community (village) surveys; baseline data are collected largely within the TL-II project and are summarized at: Dataverse and Tropical Legumes; for chickpea: Current situation; and http://exploreit.icrisat.org/. Gender data are currently being collected at 3 locations, 1 in East Africa (EA), 1 in West Africa (WA) and 1 in South Asia (SA). The data are focused on groundnut in this round of studies. Data are being collected in 2 sites each in EA and WA and in 1 site in SA. The data in North Africa (NA) are being collected for faba bean and other legumes. In SA, 1 site in Anantpur, in Andhra Pradesh; in NA, 1 site in Fez (Morocco); in EA, 1 site each in Tanzania and Uganda; in WA, 1 site each in Nigeria and Mali.	Much of the data collected on legumes are confounded by the name used for the crop and various uses of the crop. 'Bean' sometimes refers to *Phaseolus vulgaris* or *Vicia faba* and sometimes to cowpea. Data on soybean and groundnut mainly refer to these as major global commodities rather than crops for the poor, as they are for the focus of this CRP. Data for groundnut and soybean are often excluded from analyses of grain legumes even though they are major components (e.g. SPIA Report). Geographical coverage is limited for some countries, and some very important legumes such as mungbean, pea and grasspea are excluded.	Current gender data in Mali and northern Nigeria are being collected in coordination with the USAID project 'Support to Vulnerable Households.' Data sharing: PIM – shared data on baseline data, metrics and targets; Dryland Cereals – GIS-related data sets on crop locations. For common beans, a database is being established that houses metrics and indicators of interest to research and development programs. Data are standardized and will be made available to potential users. This database provides information on some of the metrics and indicators of interest to the CRPs.	Much of data collection in association with PIM and Dryland Systems. Data in relation to seed composition and bioavailability can be collected and shared with A4NH. With CCAFS, data on potential future growing areas and yield estimates. Cost-sharing will be on the basis of data use and analysis. For common bean, modifications would need to be made to the database to make it compatible with other databases and this would require additional funding. For some databases the GIS data need to be reorganized so that individuals are not easily identified. Data will be open access and freely available under a creative commons license or something similar. Data sets will include information on who collected the data, who collated or analyzed it, who was responsible for mounting the data on the internet and how these processes were funded.	Window 3 – bilateral projects contributing to the CRPs such as N2Africa, Tropical Legumes (I & II), McKnight Foundation Collaborative Crop Research Program, USAID linkage grants, the Feed the Future Innovation Labs, Indian Government and EU-IFAD, and OCP Foundation.
3.6. Dryland Cereals	CRP builds on previous work on target crops and regions. CRP uses available information on site characterization and baseline data collection from previous and ongoing work in bilateral projects within the CRP, such as Harnessing Opportunities for Productivity Enhancement of Sorghum and Millets in Sub-Saharan Africa and South Asia project, and from published literature. CRP's M&E strategy is currently being developed in detail based on a gap analysis in the context of this existing information. A baseline gender survey initiated to address 3 of its 4 crops (barley, finger millet and pearl millet) in representative sites in West & Central Africa, Eastern & Southern Africa (ESA), North Africa and South Asia. Plans to conduct a survey of various players along the entire value chain for the target crops in the target regions; will provide opportunity to identify gaps in the value chain within the regions that are critical to reach targeted goals.	R4D in crop improvement traditionally assessed by trials in research stations and on-farm trials, and this is intended to continue. Within HOPE project, HH surveys have been conducted for characterization of farmers, their trait preferences, input–output levels, and profitability of dryland cereals vis-à-vis competing crops. Instruments to be used to monitor adoption of technologies and innovations were developed and shared with economists working in the NARS institutions implementing HOPE project. 10 GIS maps had been specified within the HOPE project for Eastern and Southern Africa (ESA) in consultation with ICRISAT cereal breeders, which cover all countries in the ESA region with which ICRISAT has collaborative breeding programs for sorghum and millets. These maps are based on biophysical and socioeconomic characteristics, and 3 of the 10 GIS maps specified for the region have so far been produced. The baseline gender survey in the consultant work identified above is being based upon literature review followed by HH surveys.	Constraints include: difficulties obtaining all relevant data on smallholder agriculture; evolutions in smallholder agriculture and urbanization in certain areas, where earlier baselines might have less relevance; meager definitive data in the evolving end-use market in the various regions; time and cost involved in GIS assessment (a committed group effort for GIS mapping of common geographies for multiple CRPs will make this important activity cost efficient); insufficient support for data analysis, impact assessment (*ex ante* or *ex post*); in some instances, as in the case of barley, full picture cannot be obtained from FAOSTAT on area cropped since grain is not harvested and the crop is grazed to animals.	Entering into discussions to participate in a common effort in Burkina Faso where systems CRPs are focusing. Sharing common sites with Drylands, which offers significant potential to share and coordinate efforts, to allow crop-improvement efforts in Dryland Cereals to dovetail with crop management efforts in Dryland Systems. Joint surveys, using the same survey instruments, have been conducted for barley, potatoes and faba beans in the highlands of Ethiopia, given the geographic overlap of these crops in this region. Planned baseline work will be a collaborative effort with PIM, Dryland Systems and national partners. Currently Dryland Cereals is developing its detailed M&E strategy. Data sharing: Opportunities exist with: Drylands; WLE; CCAFS; PIM; A4NH; Livestock & Fish.	The following are being considered: joint planning meetings to define baseline data needed, overlapping regions, methods of impact assessment, responsibilities and budget; allocation of responsibilities will be based on the extent of presence and expertise in a given region for each of participating CRPs; cost-sharing will most likely be equal splitting. Data ownership will be the joint accountability of the participating CRPs and other partners.	Partners outside CGIAR will include UN Statistical Office, FAOSTAT, NARS and other national partners. Plan to identify the various baseline and impact assessment efforts ongoing in the regions, develop initial communication through the CRP and the regional offices of the Centers (ICRISAT and ICARDA), identify common interests, develop action plans and implement collaboratively. Efficiencies of data analysis can also be achieved by pooling available resources together within a given time frame to deliver timely interpretations.

CRP	Site characterization; Baseline data collection	Methods of data collection	Major gaps; Challenges	Coordination with other CRPs; Opportunities for data sharing	Data collection cost & responsibility sharing – Ownership & Attribution	Partners outside CGIAR
3.7. Livestock & Fish (L&F)	Selection of sites completed for 4 out of the 9 focal value chains. Development of IDOs indicators and targets at advanced stages. IDOs indicators have been defined and targets have been developed. IDO baselines will be collected through value chain analysis and situational analysis; we are also using project-related quantitative baselines and plan to use statistical modeling to set IDOs baselines. We are calling this whole process 'benchmarking.' It is our intention to be as systematic as possible, but this will depend somewhat on specific bilateral contexts.	Before and after (baseline and follow-up) data collection is necessary (as part of a randomized controlled trial where appropriate) to show some level of attribution toward achieving desired outcomes. Baselines will therefore be conducted for specific donor-funded projects including the eventual roll-out of integrated projects that are to be piloted during the research phase. Baseline data will be used to evaluate project results and to formulate *ex ante* impact assessments as part of our 'best bet' selection process. Conducting a representative baseline survey in order to collect IDO data for multiple value chains (i.e. for multiple countries) would be prohibitively expensive for L&F. Instead, the program will conduct a series of benchmarking exercises in which the best possible data will be used in order to establish estimated values for IDO indicators; in some cases, this may involve statistical modeling.	IDOs are very much specific to CRPs; as such there is little synergy, particularly with respect to using existing data sources and methods. Many of the indicators are highly technical and data collection methods and tools need to be developed or adapted. Setting targets has been challenging given that clear seminal sites and program timelines have not been resolved and the theories of change and impact pathways are still under development.	Had close collaboration with A4NH identifying and agreeing on nutrition and food security indicators, along with data and methods for data collection. We are planning to have similar seminal areas in countries where both L&F and A4NH operate. Even though we do not have a specific Gender IDO, we intend to adopt/harmonize gender indicators with the CGIAR Gender Network. Data sharing: There have been some joint activities with A4NH in our seminal sites, specifically with regard to disease and nutrition. We also intend to work with Humidtropics in Nicaragua and AAS in Bangladesh. Other opportunities for data sharing are still being explored.	A4NH cost sharing has been done on an ad hoc basis; for others the details around cost-sharing have not been discussed. We will follow the CGIAR Open Access Policy; otherwise will negotiate on a case-by-case basis.	Planned partnerships: 1. FAO on productivity and yield indicators and approaches for measuring greenhouse gas emissions. 2. International Labour Organization on methods for employment indicators and data. 3. IFPRI and WHO on methods and indicators of nutrition and health, particularly the computation of Dietary Diversity Index. 4. World Bank, FAO and host countries' ministries on computing and collecting data on target sub-sector share of national expenditure, policy shifts related to the target commodities, and data on target sub-sector budgeting. 5. ILRI involved in specific activities of incorporating livestock into the World Bank and FAO LSMS, which we expect to employ for our own activities.
4. Agriculture for Nutrition and Health (A4NH)	Indicators defined for all IDOs; in some cases we will be testing additional indicators as part of our research. For many of our indicators, we will use secondary data sources such as Demographic and Health Surveys (DHS), LSMS, Household Consumption and Expenditure Surveys (HCES). These are nationally representative surveys conducted regularly and data are collected at the HH and/or individual level. Baseline data on food safety IDOs are being collected with CRP L&F in their target value chains. Baseline data on policy outcomes are collected on a case-by-case basis.	Our evaluation approach is to develop impact pathways and theories of change (TOCs) for each flagship research area, and then identify the key indicators (for outcomes and assumptions/risks) along the pathway that will allow us to monitor whether the research is on track to influence IDOs. This process is under way and expected to be completed in mid-2014. While we will try to rely on secondary data for IDOs, within A4NH many types of data – individual and HH surveys, biological samples from humans and animals, on-farm trials, etc. – will be collected in the sites where we are working and used to assess progress using the TOC. Data will be collected on immediate outcomes all the way up to goal-level indicators such as stunting, micronutrient status and disease prevalence.	Have metrics for all IDO indicators but for some, such as gender, we are still doing research to determine which are best for our context. We are also looking at the 'cost-effectiveness' of indicators, e.g. where is a dietary diversity index a valid indicator? and where do we need individual food consumption data to assess diet quality and its impact on goal-level indicators like stunting? In cases such as policy, it is not the indicator as much as the approach for assessing influence on it that we are working to improve.	Worked with other CRPs on nutrition IDOs to reach agreement on the definition, indicators and metrics. We are hosting a capacity-building workshop to look at links between nutrition and gender, which will focus on impact pathways and indicators along them. We are working with CRP L&F to design, implement and analyze the baseline characterization. Data sharing: Opportunities could emerge from the coordination on the nutrition IDO and could take the form of sharing available secondary data sets and analysis – e.g. analyzing food consumption data to determine contribution of a particular food to overall dietary nutrient intake could be done more efficiently for multiple crops at once rather than crop by crop – as well as collecting primary data in shared sites. The latter is happening with L&F and opportunities may exist with others such as the systems CRPs.	This would be determined on a case-by-case basis. Follow open data policy.	We are planning to use these data. In some cases we may provide feedback to improve how data are collected.

CRP	Site characterization; Baseline data collection	Methods of data collection	Major gaps; Challenges	Coordination with other CRPs; Opportunities for data sharing	Data collection cost & responsibility sharing – Ownership & Attribution	Partners outside CGIAR
5. Water, Land and Ecosystems (WLE)	Have not started collecting baseline data. There are several reasons for this [2 pages of full detail provided in questionnaire]. Overall, at national and regional levels, baseline indicators should be the indicators of the client; WLE should piggyback on these indicators rather than collect these data itself. There is concern within WLE about 'baseline indicator collection exercise' [details in the questionnaire]. List of indicators documented in Q3; at issue is in which cases indicators would be applied and the degree to which they would be revised once research is undertaken.	WLE results framework identifies many indicators at the IDO level; these basically describe the 'resource base' (levels of desertification, salinity, irrigation and other water flows, including water accounting). Some of these can be collected through remote sensing. The theory of change is that the quality of the asset base: (1) will have a long-term impact on poverty, and (2) will reduce variability as a result of shocks (e.g. reduce income and property losses). One of the major challenges is trying to quantify trade-offs between current (higher immediate productivity) and long-term income as well as higher income versus increased stability (i.e. average income may go down, but variability may decrease). We are currently reviewing (with all 165 projects) how, in the past, they collected such data in the form of evaluations.	There are 3 related challenges: If you focus on boundary partners then you arrive at a metrics system where you 'adopt' indicators of the boundary partner. Therefore the indicators are not predefined centrally and you don't track them unless there is some evidence that the boundary partner is using your research. You work with the boundary partner to develop their capacity to collect relevant metrics and track the decisions they are making. The research, in many cases, is likely to generate the indicators (baselines). In this case as well, identifying indicators beforehand is not a productive use of resources. Without knowing boundary partners and understanding their needs, the scope of the indicators is very difficult to define.	Working with CRPs CCAFS & FTA on developing adaptive capacity indicators and frameworks. This has been a fruitful collaboration; as mentioned in the previous column, the problem is to identify the scope of these indicators and the conditions in which the framework should be applied, given the fact that the cost of collecting the indicators is likely to be prohibitive. Data sharing: See next column; whatever data we have we will share.	No concrete plans yet. Plan to undertake impact evaluations in cases where control and treatment groups can be clearly identified. This is more likely where projects are testing various 'technologies.' These outcomes are part of an assessment and learning process that can provide evidence to policy- and decision-makers interested in uptake, and are therefore considered 'very low' along the results chain. From a policy perspective impact evaluations are less likely to be possible. Outlines the reasoning behind a model of 'contribution' where baselines are collected in instances where there is uptake.	WLE scientists are participating in the development of SDGs.
6. Forests, Trees and Agroforestry (FTA)	'Sentinel Landscapes' initiative is developing a linked network of landscapes where FTA and partners collate comparative results and identify LT patterns based on a standardized set of research instruments, thus providing a framework for greater cohesion, interdependence and alignment of the operational plans across the entire FTA research portfolio, with the SRF and the development needs of partners in landscapes. FTA has set up a network of 9 sentinel landscapes in which a core set of indicators important to FTA are measured. Sampling is currently undertaken in 3 landscapes; 3 more will start early in 2014 and the other 3 are currently identifying monitoring sites and regional partners. First analytical workshop is planned for March 2014.	The baseline sampling methodology consists of 4 parts: 1. land degradation surveillance framework; 2. Village-level baselines; 3. HH surveys; 4. Tree & crop inventories on a cohort of farms from the HH survey.		FTA sentinel landscapes coincide partially with CCAFS sites (Nicaragua, Burkina Faso), WLE sentinel basins (Mekong); Humidtropics (Albertine Rift; Mekong). Data sharing: FTA data sharing platforms are: FTA Dataverse Landscape Portal; Forest Spatial Information Catalog.	FTA is in the process of drafting a theory of change for a joint research agenda and collocating research and intervention activities in Burkina Faso together with CCAFS, WLE, Drylands Burkina Faso; by March 2014 a harmonized Burkina Faso MEIA framework will be developed as a model. FTA has developed data sharing guidelines together with partners for all data collected in the sentinel landscapes. At present, metadata are shared within the network as data are collected, data are shared among pre-defined working teams and will be publicly released once they are cleaned and verified.	Core methodology has integrated metrics from already existing networks such as Poverty Environment Network (PEN), International Forest Resources and Institutions (IFRI), Centre de coopération internationale en recherche agronomique pour le développement long-term monitoring plots in production forests, livelihood metrics from the LSMS and the Grameen Progress out of Poverty, Agincourt Health Surveillance (INDEPTH Network) and the Land Degradation Surveillance Framework (LDSF). Coordination is handled by the sentinel landscape methods group.

CRP	Site characterization; Baseline data collection	Methods of data collection	Major gaps; Challenges	Coordination with other CRPs; Opportunities for data sharing	Data collection cost & responsibility sharing – Ownership & Attribution	Partners outside CGIAR
7. Climate Change, Agriculture and Food Systems (CCAFS)	Sites selected in 3 of the 5 targeted regions, and indicators have been defined and baseline data collected in all those sites. By end of 2014 baselines in the other 2 regions will be completed; we are willing to modify data collection if CRP-wide agreements have been reached by the time of baseline implementation.	We do baselines at 3 levels: HH surveys, community-level baselines (includes participatory rural appraisal), organizational baseline (surveys of key agencies involved in development that may or should have a climate lens). Metrics were designed to track food security, livelihood status, adaptation actions, mitigation actions and emissions. All available reports (and list of sites) can be found on our website and background materials (manuals, survey tools, data sets, etc.) can be found on Dataverse.	Given that IDOs are only recently being defined, we would need to revisit all the metrics in the baselines in light of the IDOs that are finally selected.	Common sites with multiple CRPs (e.g. South Asia site with 1.3, West Africa site with 1.1, 5 and 6, Latin America site with 1.2). At the Burkina Faso site there is a common monitoring plan. CRPs 5, 6 and 7 have been working together on metrics related to the adaptive capacity IDO – a report will be distributed in November 2013, and it is expected that monitoring will be started to test the framework in 2014. Data sharing: All CCAFS data are placed on Dataverse, publicly available.	Decided on a case-by-case basis. We make data publicly available through a commons agreement (on Dataverse).	Have not involved data from other agencies yet. A key partner for CCAFS baselines is University of Reading Statistical Services Centre (providing technical expertise), and data collection itself has involved a wide range of national and local partners in the target countries and regions.
8. Genebanks	Online data entry through the 'Online Reporting Tool' (ORT) for review and approval of Center genebank annual reports and workplans now in place. Baseline data collected. 50% or more of indicators and targets are elaborated. Task force formed to finalize the remaining indicators/targets.	Routine operations of genebank well documented and easily monitored through the ORT. There are published international standards for genebank operation that help guide our indicators and targets. However, the use of the collections and genebanks is poorly documented. We have to devise collective and/ or non-collective ways for ensuring that Centers gather data from users and requesters on their satisfaction with service and use of genebank materials. Various potential tools fit for different purposes: feedback forms, surveys, one-off studies, informal approaches, etc. We need to agree on methods and approaches.	Attributing impact in terms of yield, productivity, economic benefits to the service of genebanks providing the original raw materials is difficult. We would like to consider other ways of describing the unique value and service of these specific genebanks (e.g. provided).	None. Although IRRI works with breeders in a shared database to help track the use of genebank accessions in trials and crosses, etc. Data sharing: Developing a common practice whereby the use of genebank accessions within CRPs' research and breeding programs is documented and fed back to the genebanks would be useful.	N/A No plans as yet.	No plans as yet.

Annex 3. Specific metrics/indicators used/planned for monitoring progress (CRP metrics survey questionnaire October 2013)

Outcome categories	Field / Plot / Household	Village	District / Landscape	Country	Agroecology	Region
PRODUCTIVITY 1. Yield & productivity 2. Adoption	• 1, 2 **[3.1; 3.2; 3.5]**[6] Household (HH) various technologies and practices **[1.3]** • kg/ha; No. of adopters **[3.4]** • % yield increase, % profitability increase **[3.6]** • 1: annual milk yield; annual fish yield; meat yield per animal; annual kidding percentage; litter size weaning percentage; animal mortality rate • 2: adoption of new or improved technology and management practices **[3.7]** • 1 and 2: collected at field, plot, HH or individual (as in gender-disaggregated data) level, but might be analyzed at higher scales **[4]** • Net Primary Production (NPP); land use **[6]** • 1: HH productivity index; 2: No. of changes made in practices **[7]**.	• 1, 2 **[3.1; 3.2]** • 1 **[3.5]** • No. of farmer fields **[3.6]** • 2: adoption of new or improved technology and management practices **[3.7]** • NPP **[6]** • 1: community perception of change in natural resources **[7]**.	• 1 **[3.5]** • No. of farmer fields, % of female farmers **[3.6]** • Adoption of new or improved technology and management practices **[3.7]** • IRR on investments (flood harvesting, groundwater management, new irrigation schemes, RRR); % annual increase in crop yields relative to a comparable control group; No. of farmers adopting RRR (resource reuse and recycling) techniques (in millions) **[5]** • Land use, NPP **[6]**.	• 1, 2 **[3.1; 3.2]** • No. of farmer fields, % of female farmers **[3.6]** • 2: adoption of new or improved technology and management practices **[3.7]**.	• 1, 2 **[3.1; 3.2]** • No. of farmer fields, % of female farmers **[3.6]** • 2: adoption of new or improved technology and management practices **[3.7]**.	• 1, 2 **[3.1; 3.2]** • No. of farmer fields, % of female farmers **[3.6]** • 2: adoption of new or improved technology and management practices **[3.7]**.
LIVELIHOOD 1. Welfare/income 2. Food security & Nutrition	• 1, 2 **[3.1; 3.2; 3.5]** • HH income and intra-HH food and nutrition security **[1.3]** • % income increase, % increase in consumption **[3.6]** • Food security: quantity of target commodity supplied; quality of target commodity supplied; exposure to zoonotic diseases at the point of consumption; dietary diversity; consumption of target commodities by target population; level of awareness and attitudes toward dietary diversity practices by target population, level of awareness and attitudes toward consuming target commodities by target population **[3.7]** • Welfare/income: women's control of production and productive assets; total HH income; HH income from target commodity; women's control of HH income; employment opportunities; adoption of new or improved market facilities, technology and practices; diversification of products from target commodity; adoption of technologies suitable for creating employment for (low-income) women in the value chain; increased women's control of (HH) assets **[3.7]** • 1 and 2: scale as described above for productivity and adoption **[4]** • Welfare/income; food security; food composition; social networks; dependency on natural resources **[6]** • 1: sources of cash income; commercialization index; 2: No. of hunger months per year; main source of food (own farm vs. off-farm) **[7]**.	• 1, 2 **[3.1; 3.2]** • Same as for HH **[3.7]** • 2: landscape of agencies involved in food security and food crisis **[7]**.	• Same as for HH **[3.7]** • Is this where value chains fit? **[4]** • Annual % decrease in the No. of HH (including female-headed HH) below the poverty line relative to a comparable control group; No. of consumers with access to safer food (not contaminated by excreta) (in millions); % decrease in crop losses relative to a control group **[5]**.	• 1, 2 **[3.1; 3.2]** • % income increase, % increase in consumption **[3.6]** • Same as for HH **[3.7]**.		• % income increase, % increase in consumption **[3.6]** • Same as for HH **[3.7]**.

6. The numbers in bold within square brackets indicate the relevant CRPs. Key: 1.1 = Drylands; 1.2 = Humidtropics; 1.3 = AAS; 3.1 = Wheat; 3.2 = Maize; 3.3 = GRiSP; 3.4 = RTB; 3.5 = Grain Legumes; 3.6 = Dryland Cereals; 3.7 = L&F; 4 = A4NH; 5 = WLE; 6 = FTA; 7 = CCAFS.

Outcome categories	Field / Plot / Household	Village	District / Landscape	Country	Agroecology	Region
SOCIAL 1. Empowerment 2. Gender	• 2 **[3.1; 3.2.; 3.5.]** • Intra-HH empowerment **[1.3]** • % of female farmers **[3.6]** • Women's control of production and productive assess; women's control of HH income; employment opportunities; adoption of technologies suitable for creating employment for (low-income) women in the value chain; increased women's control of (HH) assets **[3.7]** • 1 and 2: scale as above **[4]** • 2: labor responsibilities on and off farm by gender **[7]**.	• 1, 2 **[3.1; 3.2; 3.5]** • Community empowerment **[1.3]** • Same as for HH **[3.7]** • Social mobility; social capital; tenure security; equity **[6]** • 2: PRA exercises completed by male/ female divided groups **[7]**	• Same as for HH **[3.7]** • Annual increase in the No. of people (including women) participating in natural resource planning and governance mechanisms **[5]** • 1: baseline looks at focus on vulnerable groups **[7]**.	• 2 **[3.1; 3.2]** • % of female farmers with yield increase, % of increased income for female farmers, processors **[3.6]** • Same as for HH **[3.7]**.		• % of female farmers with yield increase, % of increased income for female farmers, processors **[3.6]** • Same as for HH **[3.7]**.
ENVIRONMENT – NATURAL RESOURCES 1. Soil health 2. Water 3. Biodiversity 4. Climate change	• 1, 2, 3, 4 **[3.1; 3.2]** • 1, 2, 3 **[3.5]** • % reduction in area with crop failure **[3.6]** • Adoption of new technologies and management practices **[3.7]** • Soil condition; soil organic carbon; tree cover; tree diversity; herbaceous cover **[6]** • 4: Plot-level emissions (2 HH per site) **[7]**.	• Biodiversity, climate change resilience **[1.3]** • No. of HH adopting improved crop management practices **[3.6]** • Greenhouse gas emissions (GHGE); adoption of new technologies and management practices **[3.7]** • Soil condition; soil organic carbon; tree cover; tree diversity **[6]** • 1 and 2: participants reported community natural resources, including soil and water **[7]**.	• 1 **[3.5; 3.6]** • GHGE; adoption of new technologies and management practices **[3.7]** • % decrease in the No. of HH affected by drought or flood; % decrease in the area of crop affected by drought and/or flood; % annual increase in the volume of water stored as groundwater, to mitigate against drought; annual % improvement in soil degradation rates; % of land quality assessment indicators showing at least a 5% improvement **[5]** • Soil condition; soil organic carbon; tree cover; tree diversity; herbaceous cover; aboveground biomass; biodiversity; rainfall topography; wetness index **[6]**.	• % reduction in area with crop failure; No. of HH adopting improved crop management practices **[3.6]** • GHGE; adoption of new technologies and management practices **[3.7]**.	• Climate change resilience **[1.3]**.	• 1 **[3.5]** • % reduction in area with crop failure; No. of HH adopting improved crop management practices **[3.6]** • GHGE; adoption of new technologies and management practices **[3.7]**.

Outcome categories	Field / Plot / Household	Village	District / Landscape	Country	Agroecology	Region
POLICY 1. Price levels / variability 2. Policy discrimination	• 1, 2 **[3.5]** • PIM Policy Indicators **[2]**.	• 1, 2 **[3.1; 3.2]** • 1 **[3.5]**.	• Policy discrimination/progress **[1.3]** • 1 **[3.5; 3.6]** • Conducive policy and development stakeholder context in support of small-scale production and marketing systems; No. of policies affected by program interventions; No. of regulations/laws affected by program interventions; amount of private and public investment in the sub-sector; No. of partnerships established **[3.7]** • 1 and 2: in some cases other subnational level such as state or province **[4]** • No. of projects/pilots formulated as a result of WLE research; area managed under WLE inspired projects; No. of direct beneficiaries (at least 50% women) in millions; value of investments made (in projects) as a result of WLE research (US$ millions); No. of countries making significant policy changes based on WLE research; No. of countries significantly changing their resource management/ development plans based on WLE research; No. of targets in the SDGs which address key SRP issues; % of new IUCN/GEF projects using WLE assessment tools or key recommendations; WLE research cited in Ramsar declarations or manuals; No. of Global Resource Reuse Guidelines acknowledging WLE research; No. of business schools disseminating RRR technologies/ results; No. of countries where capacity development activities were implemented (outside of WLE) but based on WLE research **[5]** • 2. baseline involves CC policy progress **[7]**.	• 1, 2 **[3.1; 3.2]** • Policy discrimination/ progress **[1.3]** • 1 **[3.5]** • % reduction in price volatility **[3.6]** • Conducive policy and development stakeholder context in support of small-scale production and marketing systems; No. of policies affected by program interventions; No. of regulations/ laws affected by program interventions; amount of private and public investment in the sub-sector; No. of partnerships established **[3.7]** • 1 and 2 **[4]**.		• % reduction in price volatility **[3.6]** • Conducive policy and development stakeholder context in support of small-scale production and marketing systems; No. of policies affected by program interventions; No. of regulations/ laws affected by program interventions; amount of private and public investment in the sub-sector; No. of partnerships established **[3.7]**.

Annex 4. Estimates of the size and scale of impact in IDO indicators from the three most advanced research areas (Extracted from A4NH CRP Concept Note)

IDO: Better diet quality

Indicator	Size of impact on indicator	Scale of impact
Dietary diversity	Mean dietary diversity increased by one food group Low dietary diversity in young children (6–24 months) reduced by 10% (Integrated programs)	Will be estimated based on analysis of number of beneficiaries who could be reached through development organizations and donors that research area is seeking to influence
Intake of selected micronutrient(s) by women and children	Will be estimated by country and crop-based adoption studies, consumption rates, and results of efficacy and effectiveness studies (Biofortification) Reductions in % of mothers or young children at risk of inadequate intake of specific micronutrients (Integrated programs)	25 million micronutrient-deficient people will be reached by **biofortification** by 2018 in 8 target countries in Africa and Asia; by 2035 one billion people will have been reached Will be derived from estimates of total beneficiaries

IDO: Reduced exposure to AADs[7]

Indicator	Size of impact on indicator	Scale of impact
Exposure to pathogen/hazard in target food at point of consumption	A 10–50% reduction in exposure to pathogens common in ASF value chains among target beneficiaries by 2019 (Food safety) Target reduction in aflatoxin exposure TBD in research in Phase 1 (Food safety)	Estimate will be based on targets of CRP on Livestock and Fish (L&F) Estimated together with CRPs on Maize and Grain Legumes

IDO: Empowerment

Indicator	Size of impact on indicator	Scale of impact
Women's empowerment in agriculture index (WEAI) and other measures	To be determined. This is an active area of research to which A4NH will contribute, working with PIM and other CRPs	
Degree of participation in decisions related to food, nutrition and health	To be determined. As an example, an evaluation of a homestead food production project in Bangladesh found that the % of women who had "full participation in small household decision-making" increased from 14% to 50% (Hillenbrand, 2010)	

Better policies, programs and investments

- No. of countries that enact biofortification programs
- No. of breeding programs that include nutritional content in varietal evaluation criteria (Biofortification)
- No. of countries that integrate nutrition into their agricultural policies
- No. of NGO programs that incorporate lessons learned and findings from A4NH research into their agriculture–nutrition programming (Integrated programs)
- No. of CRPs that incorporate appropriate food safety objectives and components
- No. of countries and donors whose policies and investments in target regions support cost-effective, risk-based approaches to managing food safety.

7. AADs, aquatic animal diseases; ASF, African swine fever.

Annex 5. List of selected monitoring and metrics initiatives relevant to CGIAR mission[8]

Initiative	Organizations / Partners	Domains / Region	CGIAR-CRP Involvement / Comments
African Soil Information Service (AfSIS)	Earth Institute, CIAT-TBSF, ICRAF; ISRIC – World Soil Information, ATA Ethiopia, Kenya ARI, Malawi DARS; SARI, Tanzania	Soils; Land Degradation Surveillance Framework Sub-Saharan Africa (SSA)	**Yes** – WLE CRP, other CRPs? *Digital soil maps for SSA using new types of soil analysis and remote sensing imagery and crowdsourced ground observations; conducting agronomic field trials in selected sentinel sites.*
AGRIS – International Information System for the Agricultural Sciences and Technology	FAO	Multi-domain – Bibliographic data repositories Global	IFPRI, others? *OpenAgris is a web application that aggregates information from different web sources to provide data. Using AROVOC, it interlinks with numerous data sets (e.g. DBPedia, World Bank, Geopolitical Ontology, FAO fisheries data set, IFPRI).*
Biodiversity Information System for Europe (BISE)	EC-JRC, Eurostat, EEA	Biodiversity Indicators, Ecosystems and Natural Resources Europe	NA – *Combination of different European monitoring initiatives. A single entry point for data and information on biodiversity in the EU; serves as the EU Clearing House Mechanism to the Convention on Biological Diversity.*
CarboAfrica	Project of 6th Framework Programme of EC (UST, MPIB, ULUND, GTOS, FAO, CIRAD, CEH, CNR-IBIMET, IAO, DSA-SUN, CSIR, etc.)	Carbon fluxes and Ecosystems processes SSA	NA – *Quantification, understanding and prediction of carbon cycle and other greenhouse gases in SSA. Database and modeling approaches for up-scaling.*
CEPALSTAT	UN – ECLAC / CEPAL	Multi-domain; Economic, Socio-demographic, and Environmental LAC	NA – *Open access to more than 2000 internationally comparable statistics and indicators of Latin American and Caribbean (LAC) countries; monitoring of development outcomes and MDGs for disaggregated spatial units.*
Chinese Ecosystem Research Network (CERN)	Various Chinese institutes	Ecosystems, NRM China	NA – *Monitoring and research on ecosystems in China. Consists of 42 field research stations for various ecosystems, including agriculture, forestry, grassland and water bodies, five disciplinary centers and one synthesis center with function of data exchange and interdisciplinary research.*
CIARD Routemap to Information Nodes and Gateways (RING) for Agricultural Research for Development (ARD)	GFAR, Network participants in CIARD Movement	Agriculture, Multi-domain Global	Through GFAR *A project implemented within the Coherence in Information for Agricultural Research for Development (CIARD) initiative, facilitated by the Global Forum on Agricultural Research (GFAR). The standards implemented refer to metadata set and vocabularies, knowledge organization system/indexing scheme, format/syntax/notation, architecture/technology.*
Common Monitoring and Evaluation Framework (CMEF)	EC; European Evaluation Network for Rural Development	Rural Development, Policy EU	NA – *Provides a single framework for monitoring and evaluation of all rural development interventions.*
Committee on Sustainability Assessment (COSA)	Nonprofit global consortium of institutions (http://sustainablecommodities.org/partners)	Agriculture; Multi-domain Global	NA – *Global consortium of institutions developing and applying an independent measurement tool to analyze the distinct social, environmental and economic impacts of agricultural practices, particularly those associated with the implementation of specific sustainability programs.*
Consortium for Improving Agricultural Livelihoods in Central Africa (CIALCA)	Consortium sponsored by DGDC, Belgium and led by Bioversity, IITA and TSBF-CIAT	Multi-domain Central Africa	**Yes** *Linking 3 projects funded by the Belgian Directorate-General for Development Cooperation (DGDC), led by IITA, TSBF-CIAT and Bioversity International (2005) – Last meeting in 2011?*

8. Based on data from Shepherd *et al*. (2013); DFID-commissioned review.

Initiative	Organizations / Partners	Domains / Region	CGIAR-CRP Involvement / Comments
Demographic and Health Surveys (DHS)	ICF International, funded by USAID	Population, Health and Nutrition Global	NA – use of data *Nationally representative household surveys that provide data for a wide range of monitoring and impact evaluation indicators in the areas of population, health and nutrition.*
Famine Early Warning System (FEWSNET)	USAID, Chemonics International Inc., USGS, NASA, NOAA, USDA	Poverty and Livelihoods SSA, Central America, Afghanistan	NA – *Provider of early warning and analysis on acute food insecurity. Provides evidence-based analysis to help government decision-makers and relief agencies plan for and respond to humanitarian crises.*
GEO Global Agricultural Monitoring initiative (GEO GLAM)	GEO, University of Maryland; UMD, JRC, FAS, IRSA, CAS, ISRO, GEO Secretariat, AAFC, UCL, IIASA, USGS	Agriculture, Food Security Global	NA – *The initiative forms part of the G20 Action Plan on Food Price Volatility, which also includes the Agricultural Market Information System (AMIS, www.amis-outlook.org), another inter-institutional initiative hosted by FAO.*
GEOSHARE	Purdue University, McGill University, Stanford University, Bonn University, UKAid, IFPRI, CIAT, IRRI, CCAFS	Agriculture, Environment Global	**Yes** – several centers and CCAFS CRP *Mission to develop and maintain a freely available, global, spatially explicit database on agriculture, land use and the environment, accompanied by analysis tools and training programs.*
Global Hunger Index (GHI)	IFPRI	Food security, Agriculture Global	**Yes** – IFPRI *Designed to measure and track hunger globally and by country and region. GHI provides insights into the drivers of hunger.*
Global Information and Early Warning System (GIEWS)	FAO	Food Security, Agriculture Global	NA – data use by centers/CRPs *Tools include WinDisp: a public domain software package for the display and analysis of satellite images, maps and associated databases, with an emphasis on early warning for food security.*
Global Landscape Initiative (GLI)	Institute of Environment University of Minnesota	Agriculture, Environment Global	NA - *A program for characterizing global land use, land use changes, trends in global agricultural supply and demand, to improve ability to balance human needs with environmental stewardship, and promote secure landscapes across the globe.*
Global Terrestrial Observing System (GTOS)	UNEP, UNESCO, ICSU, FAO, WMO	Terrestrial Ecosystems Global	NA - *Program for observation, modeling and analysis of terrestrial ecosystems to support sustainable development. Facilitates access to information on terrestrial ecosystems on global and regional scales.*
Global Open Data for Agriculture and Nutrition (GODAN)	Multi-partner	Agriculture and Nutrition Global	**Yes** – CGIAR is partner *Initiative seeks to support global efforts to make agricultural and nutritionally relevant data available, accessible, and usable for unrestricted use worldwide. Focuses on building high-level policy and public and private institutional support for open data.*
Global Yield Gap and Water Productivity Atlas (GYGA)	University of Nebraska, WUR, Alterra	Agriculture Productivity (Crop production) Global	**Yes** – ICRISAT and AfricaRice are partners *A standard protocol for assessing Yp, Yw, Yg and WP is applied for all crops and countries using a bottom-up approach based on actual data and crop simulation models. Detailed maps and associated databases will be accessible through GYGA website. Aspires to global coverage of yield gaps for all major food crops and countries that produce them.*
Harvard Dataverse Network	Harvard University	Multi-domain Global	**Yes** – *Several centers and CRPs using data storage facilities Dataverse is a container for research data studies that can be customized and managed by its owner.*
HarvestChoice	IFPRI, University of Minnesota	Agriculture; Multi-domain SSA	**Yes** – several centers and CRPs involved *Spatially explicit evaluation framework to address the needs of investors, policy-makers and program managers. Online tools allow exploring data and creating maps.*

Initiative	Organizations / Partners	Domains / Region	CGIAR-CRP Involvement / Comments
Household Economy Approach (HEA)	FEG Consulting and Save the Children (UK)	Household Economy SSA, Central America, the Balkans and Asia	NA – *HEA is a livelihoods-based framework designed to provide a representation of the inside workings of household economies at different levels of a wealth continuum and in different parts of the world. Links to livelihood zones which have a strong link to agricultural activities.*
Integrated Monitoring System for African Landscapes (Vital Signs)	Conservation International, Earth Institute, CSIR	Agriculture; Ecosystem Services; Livelihood SSA	NA – *An integrated monitoring of agricultural landscapes; based on metrics gathered on the ground (household surveys), and remotely via satellites. An agricultural intensification index is proposed to combine the fraction of the landscape under transformative agricultural use with input intensification, as a fraction of the inputs required to achieve a target yield.*
Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES)	UNEP, UNDP, UNESCO, FAO	Biodiversity, Ecosystems Global	NA – *Provides assessments on biodiversity and ecosystem services. Aims to address the needs of multilateral environmental agreements related to biodiversity and ecosystem services.*
Joint Experiment for Crop Assessment and Monitoring (JECAM)	GEO Agriculture Monitoring Community of Practice – Secretariat at Agriculture and Agri-Food Canada	Agriculture; Multi-domain Global	NA – *An inter-comparison of monitoring and modeling methods, and data fusion. Data collected and shared include time series from a variety of earth observing satellites and* in situ *data.*
Land Degradation Surveillance Framework (LDSF)	ICRAF	Landscape & Ecosystems SSA	**Yes** – ICRAF, used by other CGIAR Centers *Landscape-level assessments and studies of carbon dynamics, vegetation changes, soil functional properties and soil hydrological properties.*
Landscapes for People, Food and Nature	Bioversity, ICRAF, IFAD, FAO, UNEP, WRI, CI, UNU, Netherlands, etc.	Ecosystems, Agriculture Livelihood Global	**Yes** – Bioversity, ICRAF *Mostly advocacy (no data?). Collaborative initiative to foster cross-sectoral dialogue, learning and action; aim is to understand and support integrated agricultural landscape approaches to simultaneously meet goals for food production, ecosystem health and human well-being.*
Long Term Ecological Research Network (LTER)	US Network – funded by NSF	Ecosystems; Multi-domain USA	NA – *Hosts the Network Information System (NIS) data portal; interaction with agriculture limited to one site.*
Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA)	World Bank, funded by Bill & Melinda Gates Foundation	Agriculture; Multi-domain SSA: Ethiopia, Mali, Malawi, Niger, Nigeria, Tanzania, Uganda	**Yes** – many CRPs and Centers using LSMS data *Panel household surveys with a strong focus on agriculture. Objective is to foster innovation and efficiency in statistical research on the links between agriculture and poverty reduction in the region. Explicit link between agricultural metrics and outcome metrics at the household level – but limited to 7 countries.*
Millennium Ecosystem Assessment (MA)	UN, multiple partners	World Ecosystems Global	**Yes –** part of secretariat hosted at WorldFish *Objective is to assess the consequences of ecosystem change for human well-being and the scientific basis for action needed to enhance the conservation and sustainable use of those systems. Findings provide a scientific appraisal of the condition and trends in the world's ecosystems and the services they provide (e.g. water, food, forest products, flood control and natural resources).*
National Agri-Environmental Health Analysis and Reporting Program (NAHARP)	Agriculture & Agri-Food Canada (AAFC)	Agriculture, Environment, Natural Resources Canada	NA – Provides science-based agri-environmental information to guide policy and program design. Good example of explicit link between agricultural metrics and NRM outcomes.

Initiative	Organizations / Partners	Domains / Region	CGIAR-CRP Involvement / Comments
Natural Capital Project (NatCap)	Stanford University, Nature Conservancy, WWF, University of Minnesota and global network of partners	Ecosystems, Nature Global	NA – *Provides free, open source ecosystem service software tools; e.g. InVEST, an open source software suite that enables users to quantify natural capital in biophysical, socioeconomic and other dimensions, and to visualize benefits, assess trade-offs associated with alternative choices.*
The Economics of Environmental Systems and Biodiversity (TEEB)	European Environment Agency (EEA); EU	Agriculture, Climate Change, Land Use, Biodiversity, Water Europe	NA – *An international initiative on the global economic benefits of biodiversity, loss and ecosystem degradation, drawing together expertise from the fields of science, economics and policy to enable decision-making.*
Tropical Ecology and Assessment Monitoring (TEAM)	Conservation International, Global Network	Biodiversity, Ecosystems Global (16 forest sites across Africa, Asia and Latin America)	NA – *Supporting a network of scientists committed to standardized methods of data; TEAM monitors the following metrics: terrestrial mammal and bird diversity, tree and liana diversity, aboveground carbon.*
UNEP World Conservation Monitoring Centre	UNEP-WCMC	Biodiversity and Ecosystems Global	NA – *UNEP's specialist biodiversity and assessment arm providing a range of biodiversity-related services. Sourcing, collating and sharing data information on biodiversity and ecosystems for global decision-making.*
Village Dynamics in South Asia (VDSA)	ICRISAT, IRRI, ICAR, Bill & Melinda Gates Foundation	Agriculture, rural household socioeconomics data South Asia (India and Bangladesh)	**Yes** – ICRISAT, IRRI *Project supported by the Bill & Melinda Gates Foundation, based on longitudinal Village-Level Studies of ICRISAT, provides insights into the social and economic changes in the village and household economies in the semi-arid and humid tropics. Links among livelihood outcomes, agricultural interventions and environmental drivers.*
Women's Empowerment in Agriculture Index (WEAI)	IFPRI, OPHI, USAID	Gender, Poverty and Livelihoods Global; Data sets for Bangladesh, Guatemala and Uganda	**Yes** – led by IFPRI *WEAI is a tool composed of two sub-indices: measuring how empowered women are within five domains, and gender parity in empowerment within the household. Spatial scope is still limited but implementation is expanding and provides new gender indicators.*
World Agriculture Watch (WAW)	FAO, IFAD, France, CIRAD	Agriculture, Ecosystems, Policy Global (network of local observation centers in selected representative areas)	NA – *Aims to inform policy dialogue on diversity of agricultural production systems, structural changes affecting them and implications on the 3 dimensions of sustainable development, related to key national and global challenges such as food security, poverty and NRM.*
World Database on Protected Areas	UNEP-WCMC, IUCN, WCPA, support from the private sector (Proteus Partners)	Ecosystems, Biodiversity Global	NA – *Foundation data set and maps for conservation decision-making, used for ecological gap analysis, environmental impact analysis; contains information from national governments, NGOs, academia, international and UN institutions, etc. Example of monitoring response indicator.*

Annex 6. Further examples of long-term data collection projects

In addition to long-term data collection programs by international agencies, such as the World Bank Living Standards Measurement Study (LSMS), the LSMS Integrated Surveys on Agriculture (LSMS-ISA) or the FAO Global Information and Early Warning System, there are numerous recent initiatives that could be relevant for CGIAR work on agricultural research metrics. A workshop in Rome (10–11 December 2013) brought together key players and global monitoring initiatives, including the new FAO Results Framework metrics, the LSMS-ISA, Australian Centre for International Agricultural Research (ACIAR) and EU monitoring initiatives, the World Agriculture Watch (WAW) and the Vitalsigns initiative (for details, see Workshop Report; ISPC, 2013e).

Janssen has discussed systems for monitoring, indicators and impact assessment in the EU (Janssen, 2013). Specific examples are described, including the IRENA Agri-environmental indicators used by the European Environment Agency for the integration of environmental concerns in agricultural policy, and the Common Monitoring and Evaluation Framework (CMEF), which uses about 100 indicators for evaluating rural development programs in different countries. The LIAISE (www.liaise-kit.eu/) project and network aim to link impact assessment instruments to sustainability expertise. The main lesson learned is that relevance of research for sustainable development is ensured by contextualization of research and dialogue between policy and science.

The Vitalsigns initiative aims at an integrated monitoring of agricultural landscapes in Sub-Saharan Africa (initially in Ethiopia, Ghana and Tanzania), and is led by Conservation International. The approach uses decision support indicators, which are based on consistent metrics gathered on the ground (through household surveys), and remotely via satellites. For instance, the thread for sustainable agricultural intensification aims at constructing trade-off curves; thus it has several 'top-level indices,' reflecting trade-offs among agricultural production, biodiversity, water and, potentially, social factors. The approach requires consistent measure of agricultural intensity and measures of impact. An agricultural intensification index is proposed that combines fractions of the landscape under transformative agricultural use with input intensification, as a fraction of the inputs required to achieve a target yield. The impact indicators are imported from the biodiversity, water and climate threads, and the nutrient inputs from the realized crop yield sub-thread (see http://vitalsigns.org/ for details). The World Agriculture Watch initiative is aiming to support inclusive policy dialogue on the diversity of agricultural production systems, ongoing structural changes and their impacts on key national development and global goals, such as food security and natural resource management. The WAW methodology relies on systemic and dynamic analysis of transformations, building on the sustainable livelihood framework.

Many other activities focus on metrics and indicators related to specific commodities or commodity groups. One set of activities includes the development of sustainability standards and certification systems (ISEAL Alliance, 2012). The Rainforest Alliance/Sustainable Agriculture Network (SAN) and UTZ Certified are just two examples of many organizations maintaining sustainability standards and certifying agricultural products. A direct spin-off from increased product certification is the greater availability of consumer information on the sustainability of products and brands. Besides labeling, good examples of indicator systems include Oxfam's 'Behind the Brands' campaign (www.behindthebrands.org), the Good Guide (www.goodguide.com) and the Ecolabel Index (www.ecolabelindex.com). The last 10 years have seen a proliferation of certification systems.

Recently, there has been an increasing involvement by industry (e.g. producers, processors, traders and finance providers) and the creation of multistakeholder platforms (van Dam *et al*., 2006; Zarilli, 2008). Sustainability standards and certification systems for a number of global commodity chains have been the achievement of these multistakeholder platforms (Markevičius *et al*., 2010). Two examples of these multistakeholder platforms that bring together industry, civil society and research are the Roundtable for Sustainable Palm Oil (RSPO) and the Roundtable on Responsible Soy (RTRS). Consistently, the standards and schemes developed by such roundtables are underpinned by hierarchical frameworks of principles, criteria and indicators (RSB, 2011; RSPO, 2013).

Annex 7. Common IDOs among the CRPs – Links with SLOs taken from the common IDO table in the April 2014 guidelines

Common IDO	CRP IDO
1. **Productivity** – Improved productivity in low-income food systems Contributing mainly to **SLO2 Food Security** **10 CRPs involved** **14 IDOs**	*1.2 Humidtropics IDO 3:* Increased total factor productivity of integrated systems *1.3 AAS IDO3:* Improved productivity in aquatic agricultural systems (water and total factor productivity) *3.1 WHEAT IDO4:* Smallholders' use of modern wheat varieties translates into higher, more stable yields in WHEAT target regions *3.1 WHEAT IDO5:* Faster and more significant genetic gains in better breeding programs worldwide, using more effective approaches for complex traits *3.2 Maize IDO1:* Increased productivity and stability of farming systems *3.2 Maize IDO3:* Increased yields of maize for smallholder farmers *3.2 Maize IDO5:* Reduced postharvest losses *3.3 GRiSP IDO1:* Increased rice yield *3.3 GRiSP IDO2:* Increased rice productivity (or resource-use efficiency) *3.4 RTB IDO1:* Improved productivity in smallholder RTB cropping systems *3.5 Grain Legumes IDO4:* Improved productivity of farming systems, especially among smallholder farmers *3.6 Dryland Cereals IDO1:* Improved productivity of dryland cereals in smallholder farming in Africa and Asia *3.7 L&F IDO1:* Increased livestock and fish productivity in small-scale production systems for the target commodities *5 WLE IDO1:* Sustainable increases in water, land and energy productivity in rainfed and irrigated agroecosystems
2. **Food security** – Increased and stable access to food commodities by rural and urban poor Contributing mainly to **SLO2 Food Security** **6 CRPs involved** **6 IDOs**	*3.4 RTB IDO2:* Increased and stable access to food commodities by rural and urban poor *3.5 Grain Legumes IDO1:* Improved and stable access to grain legumes by urban and rural poor *3.6 Dryland Cereals IDO2:* Increased and stable access to dryland cereal food, feed and fodder by the poor, especially rural women and children *3.7 L&F IDO2:* Increased quantity and improved quality of the target commodity supplied from the target small-scale production and marketing systems *6 FTA IDO5:* Production and availability of foods, fuel and other products from FTA systems increased for poor dependent people *7 CCAFS IDO1:* Increased and stable access to food commodities by rural and urban poor
3. **Nutrition** – Improved diet quality of nutritionally vulnerable populations, especially women and children Contributing mainly to **SLO3 Nutrition & Health** **10 CRPs involved** **11 IDOs**	*1.1 Dryland Systems IDO3:* Women and children in vulnerable households have year-round access to greater quantity and diversity of food sources *1.2 Humidtropics IDO2:* Increased consumption of safe, nutritious foods by the poor, especially among nutritionally vulnerable women and children *1.3 AAS IDO2:* Increased consumption of nutritious, safe foods by low-income households in aquatic agricultural systems, especially by nutritionally vulnerable women and children *3.2 Maize IDO4:* Increased nutritional diet *3.3 GRiSP IDO6:* Improved nutrition status derived from rice consumption *3.4 RTB IDO4:* Increased consumption of safe and nutritious food by the poor especially among the nutritionally vulnerable women and children *3.5 Grain Legumes IDO3:* Increased consumption of healthy grain legumes and products by the poor for a more balanced and nutritious diet, especially among nutritionally vulnerable women and children *3.6 Dryland Cereals IDO3:* Increased consumption of nutritious dryland cereals by the poor, especially rural women and children *3.7 L&F IDO4:* Increased consumption of the target commodity responsible for filling a larger share of the nutrient gap for the poor, particularly for nutritionally vulnerable populations (women of reproductive age and young children) *4 A4NH IDO1:* Better diet quality
4. **Income** – Increased and more equitable income from agricultural and natural resources management and environmental services earned by low-income value chain actors Contributing mainly to **SLO1 Poverty Reduction** **SLO2 Food Security** **SLO3 Nutrition & Health** **11 CRPs involved** **11 IDOs**	*1.1 Dryland Systems IDO2:* More stable and higher per-capita income for 'intensifiable households' *1.2 Humidtropics IDO1:* Increased and more equitable income from agriculture for rural poor farm families, with special focus on rural women *1.3 AAS IDO1:* Increased and more equitable income from agricultural and natural resource management and environmental services earned by low-income value chain actors in aquatic agricultural systems *3.2 Maize IDO2:* Increased and more equitable income for men and women smallholder farmers from adopting improved maize varieties *3.3 GRiSP IDO3:* Decreased poverty of net rice consumers (urban and rural) and rice producers *3.4 RTB IDO3:* Increased and more gender-equitable income for poor participants in RTB value chains *3.5 Grain Legumes IDO2:* Increased and more equitable income from grain legumes by low-income value chain actors, especially women *3.6 Dryland Cereals IDO4:* Increased and more equitable income from marketing dryland cereal grain, fodder and products by low-income value chain actors, especially smallholder women *3.7 L&F IDO3:* Increased employment and income for low-income actors in the target value chains, with an increased share of employment for and income controlled by low-income women *5 WLE IDO2:* Increased and more equitable income from agricultural and natural resources management and ecosystem services in rural and peri-urban areas *6 FTA IDO4:* Income from products and environmental services derived from forests, trees and agroforestry systems enhanced

Common IDO	CRP IDO
5. **Gender & Empowerment** – Increased control over resources and participation in decision-making by women and other marginalized groups Contributing mainly to **SLO1 Poverty Reduction** **7 CRPs involved** **7 IDOs**	*1.2 Humidtropics IDO5:* Increased control by women and other marginalized groups over integrated systems assets, inputs, decision-making and benefits *1.3 AAS IDO4:* Increased control of assets, inputs, decision-making and benefits by women and other marginalized groups in aquatic agricultural systems *3.3 GRiSP IDO9:* Increased gender equity in the rice value chain *4 A4NH IDO3:* Empowerment *5 WLE IDO3:* Women and marginalized groups have improved decision-making power over and increased benefits derived from agriculture and natural resources *6 FTA IDO3:* Greater gender equity in decision-making and control over forest and tree use, management and benefits are improved through women's empowerment *7 CCAFS IDO2:* Increased control by women and other marginalized groups of assets, inputs, decision-making and benefits
6. **Capacity to Innovate** – Increased capacity for innovation within low-income and vulnerable rural communities allowing them to improve livelihoods Contributing mainly to **SLO1 Poverty Reduction** **SLO2 Food Security** **SLO3 Nutrition & Health** **2 CRPs involved** **2 IDOs**	*1.2 Humidtropics IDO6:* Increased capacity for integrated systems to innovate and bring social and technical solutions to scale *1.3 AAS IDO5:* Increased capacity to innovate within low-income and vulnerable rural communities in aquatic agricultural systems allowing them to seize new opportunities to improve livelihoods and increase household income
7. **Adaptive capacity** – Increased capacity in low-income communities to adapt to environmental and economic variability, shocks and longer term changes Contributing mainly to **SLO 1 Poverty Reduction** **SLO 2 Food Security** **6 CRPs involved** **6 IDOs**	*1.1 Dryland Systems IDO1:* More resilient livelihoods for vulnerable households in marginal areas *1.3 AAS IDO6:* Increased capacity to adapt to environmental and economic variability, shocks and longer term changes in low-income communities in aquatic agricultural systems *3.6 Dryland Cereals IDO5:* Increased capacity to adapt to environmental variability and longer term changes in low-income communities in Africa and Asia *5 WLE IDO4:* Increased ability of low-income communities to adapt to environmental and economic variability, demographic shifts, shocks and long-term changes *6 FTA IDO6:* Resilience to environmental and economic variability, shocks and longer term changes of rural communities enhanced through greater adaptive capacity to manage FTA systems *7 CCAFS IDO3:* Increased capacity in low-income communities to adapt to climate variability, shocks and longer term changes
8. **Policies** – More effective policies and institutions supporting sustainable, resilient and equitable agricultural and natural resources management developed and adopted by agricultural, conservation and development organizations, national governments and international bodies Contributing mainly to **SLO1 Poverty Reduction** **SLO2 Food Security** **8 CRPs involved** **14 IDOs**	*1.1 Dryland Systems IDO7:* Policy reform removing constraints and creating incentives for rural households to engage in more sustainable practices that improve resilience and intensify production *2 PIM IDO1:* Improved prioritization of global agricultural research effort for developing countries *2 PIM IDO2:* In selected countries of focus, attainment of target levels of investment in agricultural research and rates of return to research that at least meet global averages *2 PIM IDO3:* Increased adoption of superior technologies and management practices in relevant domains of application *2 PIM IDO4:* Improved sectoral policy and better public spending for agriculture in agriculturally dependent developing countries *2 PIM IDO5:* Strengthened value chains that link producers and consumers with lower transactions costs, increased inclusion of smallholders, and provision of benefits to both women and men *2 PIM IDO6:* Improved design and coverage of social protection programs with particular emphasis on vulnerable rural populations *2 PIM IDO7:* Improved use of scientific evidence in decision processes related to sustainability of natural resources important for rural livelihoods *3.4 RTB IDO7:* Enabling policy environment supporting development and use of pro-poor and gender-inclusive RTB technologies *3.7 L&F IDO6:* Policies (including investments) support the development of small-scale production and marketing systems, and seek to increase the participation of women within these value chains *4 A4NH IDO4:* Better policies, programs and investments *6 FTA IDO1:* Policies and practices supporting sustainable and equitable management of forests and trees developed and adopted by conservation and development organizations, national governments and international bodies *6 FTA IDO2:* Local institutions strengthened and collective action enhanced for improved forest and tree management in landscapes *7 CCAFS IDO4:* Additional policies and institutions supporting sustainable, resilient and equitable agricultural and natural resources management developed and adopted by agricultural, conservation and development organizations, national governments and international bodies

Common IDO	CRP IDO
9. **Environment** – Minimized adverse environmental effects of increased production intensification Contributing mainly to **SLO4 Sustainability** **5 CRPs involved** **6 IDOs**	*1.2 Humidtropics IDO4:* Reduced adverse environmental effects of integrated systems intensification and diversification *3.1 WHEAT IDO2:* Farmers minimize unsustainable effects on soil, environment and improve their household income and livelihoods *3.3 GRiSP IDO2:* Increased rice productivity (or resource-use efficiency) *3.3 GRiSP IDO4:* Increased sustainability and environmental quality of rice-based cropping systems *3.4 RTB IDO5:* Minimized adverse environmental effects of increased RTB production, processing and intensification *3.5 Grain Legumes IDO5:* Minimized adverse environmental effects of increased production and intensification of grain legumes
10. **Future Options** – Greater resilience of agricultural/ forest/ water-based/mixed crop livestock, aquatic systems for enhanced ecosystem Contributing mainly to **SLO2 Food Security** **SLO4 Sustainability** **4 CRPs involved** **4 IDOs**	*3.3 GRiSP IDO7:* Increased rice genetic diversity for current and future generations *3.4 RTB IDO6:* Improved ecosystem services for enhanced food system stability and sustaining novel genetic diversity for future use *5 WLE IDO5:* Increased resilience of communities through enhanced ecosystem services in agricultural landscapes *6 FTA IDO7:* Biodiversity and ecosystem services (including carbon sequestration) from forests and trees conserved or improved in key target countries
11. **Climate** – Increased carbon sequestration and reduction of greenhouse gases through improved agriculture and natural resources management Contributing mainly to **SLO4 Sustainability** **2 CRPs** **2 IDOs**	*3.7 L&F IDO5:* Lower environmental impacts in the target value chains *7 CCAFS IDO5:* Increased carbon sequestration and reduction of greenhouse gases through improved agriculture and natural resources management
CRP-specific IDOs	
1.1 Dryland Systems	*IDO4:* More sustainable and equitable management of land and water resources in pastoral and agropastoral systems *IDO5:* Better functioning markets underpinning intensification of rural livelihoods *IDO6:* More integrated, effective and connected service delivery institutions underpinning resilience and system intensification
3.1 WHEAT	*IDO3:* Farmers have more and better access to quality seeds and use them
3.2 Maize	*IDO6:* Reduced aflatoxin in maize value chain
3.3 GRiSP	*IDO5:* Improved efficiency and increased value in rice value chain *IDO8:* Increased pro-poor and gender-equitable delivery systems for improved rice technologies
4 A4NH	*IDO2:* Reduced exposure to agriculture-related diseases

Annex 8. Previous concerns over data archiving and storage in CGIAR

> "The summary report from the June 2008 CGIAR Data Management Workshop hosted by Bioversity in Rome sums it up this way in its opening sentences: 'Empirical data from field and lab observation are, when connected with secondary information, the raw material of all our research outputs. They must therefore be valuable, yet we look after them in a surprisingly casual way. Compared with our financial data, we have few standards or recognized good practices, few professional staff responsible, few incentives for good performance and no indicators of success.' The Panel completely agrees."
>
> —Stripe Review of the Social Sciences in the CGIAR (ISPC, 2009)

Data are still viewed as the property of individual scientists, an antiquated approach that invites unnecessary waste and does not safeguard against scientific fraud by limiting opportunities for replication. Inter-Center cooperation on data management issues is surprisingly limited although there are enormous economies of scale to be had in data management. The CGIAR Information and Communications Technology – Knowledge Management (ICT-KM) Program (http://ictkm.cgiar.org/) is attentive to this issue but progress to date has been minimal.

Outside of the HarvestChoice project, there has been no significant investment in creating metadata that might make publicly available the data sets that CGIAR scientists invest their scarce time and skill in collecting. Making data available helps investigators in other research institutions make more and greater discoveries and with less duplication. Data availability is therefore an important international public good. It requires adequate and careful documentation of the survey methods and instruments used, data cleaning and variable construction procedures, etc., so that the data are properly contextualized and interpreted. Although there are clearly major economies of scale and scope involved in metadata creation, maintenance and dissemination, we saw little evidence of substantive system-wide efforts in this direction, whether through the CGIAR ICT-KM Program or the Consortium for Spatial Information, either of which could, in principle, make important advances in this area with the right leadership and incentives.

> "… the CGIAR has paid insufficient attention to maintaining a central facility for collection, quality control and archiving of data. While experimental data from research conducted on the main research stations may be archived, the stored files often lack sufficient meta-data and annotation to allow their ready use. The situation concerning data from experiments conducted on smaller research stations or in farmers' fields is parlous. Data have often disappeared with departing scientists or have been lost due to problems with disk storage. The conduct of syntheses or follow-up studies is thus compromised. A promising initiative established through the CGIAR Research Programme on Climate Change, Agriculture and Food Security (CCAFS) is the Global Agricultural Trial Repository. At the time of writing (mid-2012) some 2500 trials had been uploaded. The Panel therefore recommends that all programmes be charged with contributing their data to this, or similar repositories agreed at the Consortium level. Programmes will have to ensure that data are developed and stored as more comprehensive meta-data sets for comparability across programmes. Thus, they can serve collaborative research and learning, as well as opportunities to enhance the development of international public goods."
>
> —Stripe Review of Natural Resources Management Research in the CGIAR (ISPC, 2012a)

Annex 9. Tentative alignment of CGIAR SLOs, the common set of CRP IDOs and Sustainable Development Goals (SDGs) and a critique of potential indicators

SDG[9]	CGIAR SLOs	Common set of CRP IDOs	SDG targets and indicators (examples)
1. End extreme poverty including hunger	SLO1 – Poverty reduction SLO2 – Food security SLO3 – Nutrition & health	Food security – *Increased and stable access to food commodities by rural and urban poor* Income – *Increased and more equitable income from agricultural and natural resources management (NRM) and environmental services earned by low-income value chain actors* Nutrition – *Improved diet quality of nutritionally vulnerable populations, especially women and children*	**Target 01a.** End extreme poverty, including absolute income poverty ($1.25 or less per day). Indicators: • Proportion of population below $1.25 (PPP) per day (MDG indicator) (SDG indicator no. 1) • [Proportion of population in extreme multidimensional poverty – *Indicator to be developed*] (2) **Target 01b.** End hunger and achieve food security, appropriate nutrition, and zero child stunting. Indicators: • Prevalence of stunting in children under 5 years of age (3) • Proportion of population below minimum level of dietary energy consumption (%) (4) • [Proportion of population with shortfalls of any one of the following essential micronutrients: iron, zinc, iodine, vitamin A, folate and vitamin B12 – *indicator to be developed*] (5)
4. Achieve gender equality, social inclusion and human rights for all	SLO1 – Poverty reduction	Gender & empowerment – *Increased control over resources and participation in decision-making by women and other marginalized groups*	**Target 04b.** Reduce by half the proportion of households with incomes less than half of the national median income (relative poverty). Indicators: • Proportion of households with incomes below 50% of median income (relative poverty) (30) • Gini coefficient (31)
5. Achieve health and well-being at all ages	SLO3 – Nutrition & health	Nutrition – *Improved diet quality of nutritionally vulnerable populations, especially women and children*	**Target 05c.** Implement policies to promote and monitor healthy diets, physical activity and subjective well-being, etc. Indicators: • Household Dietary Diversity Score (46)
6. Improve agricultural systems and raise rural prosperity	SLO2 – Food security SLO1 – Poverty reduction	Productivity – *Improved productivity in pro-poor food systems* Adaptive capacity – *Increased capacity in low-income communities to adapt to environmental and economic variability, shocks and longer term changes*	**Target 06a.** Ensure sustainable food production systems with high yields and high efficiency of water, nutrients and energy, supporting nutritious diets with low food losses and waste. Indicators: • Crop yield gap (actual yield as % of attainable yield) (50) • Crop nitrogen-use efficiency (%) (51) • [Crop water productivity (tons of harvested product per unit irrigation water) – *indicator to be developed*] (52) • [Share of agricultural produce loss and food waste (% of food production) – *indicator to be developed*] (53) **Target 06b.** Halt forest and wetland conversion to agriculture, protect soil and land resources, and ensure that farming systems are resilient to climate change and disasters. Indicators: • Annual change in forest area and land under cultivation (54) • Annual change in degraded or desertified arable land (% or ha) (55) • Economic losses from disasters in rural areas due to climatic and non-climatic events (in US$) [*indicator to be specified*] (56) **Target 06c.** Ensure universal access in rural areas to basic resources and infrastructure services (land, water, sanitation, modern energy, transport, mobile and broadband communication, agricultural inputs and advisory services). Indicators: • Percentage of rural population using basic drinking water (57) • Percentage of rural population using basic sanitation (58) • [Access to drying, storage and processing facilities – *indicator to be developed*] (61) • [Share of farmers covered by agricultural extension or equivalent programs – *indicator to be developed*] (62)

9. SDSN (2014).

SDG[9]	CGIAR SLOs	Common set of CRP IDOs	SDG targets and indicators (examples)
8. Curb human-induced climate change and ensure sustainable energy	SLO4 – Sustainability	Climate – *Increased carbon sequestration and reduction of greenhouse gases through improved agriculture and NRM*	**Target 08b.** Reduce non-energy-related emissions of greenhouse gases (GHGs) through improved practices in agriculture, forestry, waste management and industry. Indicators: • Net GHG emissions in the agriculture, forest and other land use (AFOLU) sector (tCO_2e) (77)
9. Secure ecosystem services and biodiversity, and ensure good management of water and other natural resources	SLO4 – Sustainability	Environment – *Minimized adverse environmental effects of increased production intensification*	**Target 09a.** Secure ecosystem services by adopting policies and legislation that address drivers of ecosystem degradation, and requiring individuals, businesses and governments to pay the social cost of pollution and the use of environmental services. Indicators: • Red List Index (Biodiversity – by country and major species group) (80) • Area of forest under sustainable forest management (%) (82) **Target 09b.** Participate in and support regional and global arrangements to inventory, monitor and protect ecosystem services and environmental commons of regional and global significance and curb trans-boundary environmental harms, with robust systems in place no later than 2020. Indicators: • Proportion of fish stocks within safe biological limits (83) • Red List Index (Biodiversity – for Internationally Traded Species) (80) **Target 09c.** All governments and businesses commit to the sustainable, integrated and transparent management of water, agricultural land, forests, fisheries, mining and hydrocarbon resources to support inclusive economic development and the achievement of all SDGs. Indicators: • Proportion of total water resources used (MDG indicator) (85) • Access to land in rural areas index (86)
10. Transform governance for sustainable development	All 4 SLOs	Policies – *More effective policies, supporting sustainable, resilient and equitable agricultural and NRM developed and adopted by agricultural, conservation and development organizations, national governments and international bodies* Capacity to innovate – *Increased capacity for innovation within low-income and vulnerable rural communities allowing them to improve livelihoods* Future options – *Greater resilience of agricultural/ forest/water-based/mixed crop livestock, aquatic systems for enhanced ecosystem services*	**Target 10b.** Adequate domestic and international public finance for ending extreme poverty, providing global public goods, capacity building and transferring technologies, including 0.7% of GNI in ODA for all high-income countries, and an additional $100 billion per year in official climate financing by 2020. Indicators: • Domestic revenues allocated to sustainable development as % of GNI (94) **Target 10c.** Accelerate adoption of new technologies for the SDGs. Indicators: • Researchers and technicians in R&D (per million people) (100)

11. SDSN (2014).

Independent
Science and
Partnership
Council

CGIAR Independent Science & Partnership Council (ISPC) Secretariat
c/o FAO, Viale delle Terme di Caracalla
00153 Rome, Italy
t: +39 06 570 52103
ISPC-Secretariat@fao.org / http://ispc.cgiar.org